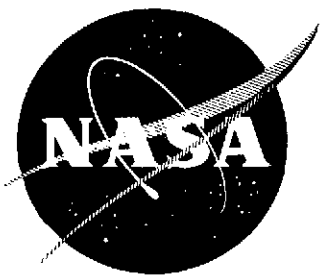
NASA CR-140202
ERIM 190100-31-T

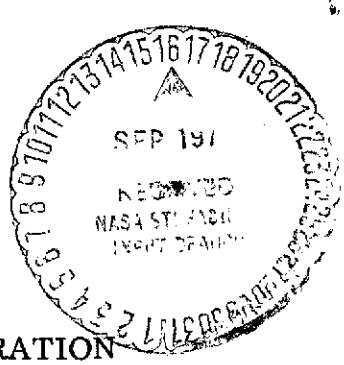


ADAPTIVE PROCESSING WITH A DECISION-DIRECTED KALMAN FILTER AND FEATURE EXTRACTION OF MULTISPECTRAL DATA

by

R. B. Crane
Infrared and Optics Division

July 1974
prepared for



NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Johnson Space Center, Houston, Texas 77058
Earth Observations Division
Contract NAS 9-9784, Task VIII

NOTICES

Sponsorship. The work reported herein was conducted by the Environmental Research Institute of Michigan for the National Aeronautics and Space Administration, Johnson Space Center, Houston, Texas 77058, under Contract NAS 9-9784, Task VIII. Dr. Andrew Potter/TF3 is Technical Monitor. Contracts and grants to the Institute for the support of sponsored research are administered through the Office of Contracts Administration.

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. NASA CR-ERIM 190100-31-T	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle ADAPTIVE PROCESSING WITH A DECISION-DIRECTED KALMAN FILTER AND FEATURE EXTRACTION OF MULTISPECTRAL DATA		5. Report Date December 1973 6. Performing Organization Code
7. Author(s) Robert B. Crane		8. Performing Organization Report No. ERIM 190100-31-T
9. Performing Organization Name and Address Environmental Research Institute of Michigan Infrared and Optics Division P. O. Box 618 Ann Arbor, Michigan 48107		10. Work Unit No. Task VIII 11. Contract or Grant No. NAS 9-9784
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Johnson Space Center Earth Observations Division Houston, Texas 77058		13. Type of Report and Period Covered Technical Report 1 February 1973 through 31 October 1973 14. Sponsoring Agency Code

15. Supplementary Notes

Dr. Andrew Potter/TF3 is Technical Monitor for NASA.

16. Abstract

A class of adaptive decision-directed classifiers based on the Kalman filter theory are presented. Test results are shown for some simple classifiers which indicate that adapting improves classification accuracy. Memory and computational requirements can be reduced significantly over those normally associated with Kalman filters. Auxiliary training data can be used in a way that should reduce the probability of signature capture.

A method is presented for feature extraction of multispectral scanner data. The method is the formation of a subset of linear combinations of the data from all the spectral channels prior to recognition processing. This method is compared to the selection of a subset of the pure spectral channels prior to recognition processing. Training data are used to define both processes. Non-training data are then used to demonstrate the level of performance, given the reduction in the number of channels entering the recognition processor and using feature extraction rather than feature selection.

17. Key Words	18. Distribution Statement
Decision-directed classifiers Kalman filter Multispectral scanner data Signature capture	Initial distribution is listed at the end of this document.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
UNCLASSIFIED	UNCLASSIFIED	63	

PREFACE

This report describes part of a comprehensive and continuing program of research concerned with advancing the state-of-the-art in remote sensing of the environment from aircraft and satellites. The research is being carried out for the NASA Lyndon B. Johnson Space Center, Houston, Texas, by the Environmental Research Institute of Michigan (ERIM), formerly the Willow Run Laboratories of The University of Michigan. The basic objective of this multidisciplinary program is to develop remote sensing as a practical tool to provide the planner and decision-maker with extensive information quickly and economically.

Timely information obtained by remote sensing can be important to such people as the farmer, the city planner, the conservationist, and others concerned with problems such as crop yield and disease, urban land studies and development, water pollution, and forest management. The scope of our program includes: (1) extending the understanding of basic processes; (2) discovering new applications, developing advanced remote-sensing systems, and improving automatic data processing to extract information in a useful form; and also (3) assisting in data collection, processing, analysis, and ground-truth verification.

The research described here was performed under NASA Contract NAS 9-9784, Task VIII, and covers the period from 1 February 1973 through 31 October 1973, Dr. Andrew Potter has been Technical Monitor. The program was directed by R. R. Legault, a Vice-President of ERIM, by J. D. Erickson, Principal Investigator and Head of the Information Systems and Analysis Department, and by R. F. Nalepka, Head of the Multispectral Analysis Section. The ERIM number for this report is 190100-31-T.

The author wishes to acknowledge the direction provided by Mr. R. R. Legault, Dr. J. D. Erickson, and Mr. R. F. Nalepka. A. Sato and J. F. Reyer expertly translated equations into computer language and obtained tests results. Many constructive discussions were held with the above and with Prof. W. L. Root, H. Horwitz and R. J. Kauth.

CONTENTS

FIGURES

TABLES

1
SUMMARY

A general form of a Kalman filter model is presented and shown to be applicable to an adaptive method of classification of remotely sensed data. The mean vectors for all of the signatures can be adapted, using a decision-directed Kalman filter and known additive and multiplicative changes in the mean vectors. A simplification of the general form, which updates the estimates of the mean vectors after each decision, is developed and tested. Previously used methods incorporating interactive and non-interactive updating are shown to be simple Kalman filters.

A processing method is shown that reduces memory and computation requirements over those normally associated with a Kalman filter. A formulation of the Kalman filter is presented that updates the mean vectors after each line, rather than after each data point. Included in the formulation is the use of a factor that depends on the degree of confidence that the decision made for each data point is correct. Methods are suggested for adapting the preprocessing functions (scan-angle correction) and for using auxiliary ground truth data to reduce the probability of signature capture.

Test results are presented which tend to confirm the usefulness of the Kalman filter. It is shown that classification accuracy is dependent on certain filter parameters, the parameters which determine updating rate, interaction, etc. We also show that there is essentially no change in classification accuracy when the updating is computed after every line of data, rather than after each data point.

There appear to be two ways in which the Kalman filter increases classification accuracy. One way is to update the means to follow rapid changes in the data, such as might occur in data from adjacent fields of the same class. The other way is to update at a slower rate so that the updated means adjust for data changes such as those caused by atmospheric changes.

Feature extraction, whereby the dimensionality of the multispectral data is reduced in such a manner that there is no significant loss in classification accuracy, is also discussed. A previously used method, feature selection (use of a subset of spectral channels), is one way to implement feature extraction. A greater reduction in dimensionality may be obtained by use of a subset of linear combinations.

The general problem of finding a suitable subset of linear combinations is included here. Part of the general problem is that of selecting a distance measure, and one particular distance measure is recommended because it is more closely related to classification accuracy than some of the more popular distance measures.

Limited test results show that a subset of 3 linear combinations can produce a higher classification accuracy than a subset of 4 spectral channels.

2
INTRODUCTION

Multispectral scanners, carried aboard aircraft or spacecraft, are being used to survey a wide range of earth's resource and environmental parameters. The Environmental Research Institute of Michigan (ERIM) is engaged in a broad-based research effort, developing and applying practical techniques of extracting information from the data generated by these instruments. Considering the spacecraft case for example, the scanner field of view sweeps laterally across the terrain, and, as the spacecraft moves forward, successive scans completely cover the scene below. The spectrum of solar radiation reflected by the terrain is sensed by the scanner. In particular, the scanner records, for each resolution element in the scene, a data vector (signal) whose elements are proportional to the radiance at each of several wavelengths in the spectrum.

The processing of these data to extract information from the scene may be considered under two broad categories: classification procedures and estimation procedures. In estimation, we attempt to answer questions of temperature, moisture, or biomass for each resolution element. In classification, we attempt to decide for each resolution element (data vector) what is the class of material present. The remainder of this report is concerned with classification procedures, and to focus on a specific problem, with the classification of major agricultural crop species such as corn, wheat, and soybeans. We wish to focus attention on an adaptive classifier technique based on the Kalman filter although we also discuss some related topics such as feature selection and preprocessing.

We have found through experience that, for a homogeneous class or subclass of materials on the ground, a Gaussian distribution is a useful characterization of the distribution of data points from that class of material, so that the mean vector and the covariance matrix are sufficient parameters to use as input to a classifier [1]. We have used a variety of classifiers—maximum likelihood, linear decision trees, shortest distance, etc.—based on this mean and covariance description for each class or subclass of material.

We have also found through experience that there are variations in the mean and covariance matrix as we consider data from different flights, and even from different fields of the same class near each other [2]. These variations of means and covariance matrices are due to changes in sun angle, amounts of haze, cloud and cloud shadows, moisture and fertilizer content of soil, variations in planting time and development rate for crops of the same type, and differences in planting density, to name most of the causative factors. One variation which is always present is the dependence of the reflectance of the crop and of the transmission and scattering effects of the atmosphere on angle of view (scan angle). Thus, the mean and covariance of the signal from a particular crop tend to be a function of scan angle even within a particular homogeneous field.

Because these variations reduce classification accuracy, several parallel research efforts are being carried out at ERIM to develop methods to compensate for the variations in the means and covariances. One approach is based on in-scene references where changes that occur in selected fields are assumed to reflect changes in all of the data [3]. A second approach uses functions of the data, such as ratios of channels, for classification [4-7]. Generally, functions cannot be found that are completely independent of atmospheric or ground cover changes. The third class uses radiance and reflectance models to estimate and account for the changes that are occurring [8-10].

The method being discussed here is that of making adaptive corrections [11, 12]. In this approach, the mean vectors are slowly updated based on the decisions made by the classifier and on the actual values of the individual data vectors which are classified. The approach is based on the following idea. Suppose a sequence of observations (data vectors), z_j, z_{j+1}, . . . were all recognized as material class A by the classifier, but that these observations tended to cluster to one side of the current estimate of the mean, μ_A, of that material class. This would provide us with some evidence that the mean of the material class A had shifted. A decision-directed adaptive classifier is one which automatically adjusts the value of μ_A so as to bring it closer to the current observations which were classified as material A.

We would like our decision-directed adaptive classifier to take account of some additional considerations. The amount by which we allow a signature to be modified in any particular updating cycle may be different in different spectral channels. Also, a particular crop may not be observed for some time, and during that time the true mean of that crop, along with the means of other crops, may shift. Hence we would like to be able to adapt all signatures based upon the observations and classifications of one or a few of them.

In practice, resolution elements often overlap two or more different crop types, producing an observation far from the mean of any particular crop class. We would like to avoid using these observations as well as "wild" observations from any other cause.

Kalman filter theory provides a framework within which these considerations and others can be combined into one systematic approach [12, 13].

The Kalman filter is an iterative filter, especially useful for digital computation, that produces an estimate of a time sequence of state vectors from a corresponding time sequence of measurement vectors. In the simplest application,five elements must be defined. These are: (1) the state vector, (2) the measurement vector, (3) an observation matrix relating the state vector to the measurement vector (assuming no measurement noise) by a linear transformation, (4) a covariance matrix describing additive noise in the measurement, and (5) a covariance matrix describing the statistics of the successive differences in the state vector.

In order to apply the Kalman filter to remote sensing data, we must make an association between the elements of the Kalman filter and elements of the classifier. This can be done in a number of ways, one of which is now described.

Assume that the most important statistics to update are the components of the mean vector of each material class, and that we will update after each single observation. Then we make the following identifications.

(1) The mean vectors of each material are combined into a single vector identified as the state vector, x_t. The initial condition, x_0, is given by the initial training data for each crop.

(2) The observed data vector is identified as the measurement, z_t.

(3) The classified output (a recognition vector) is used to produce a matrix, H_t, of zeros and ones (a spotting function) which selects the correct components of the state vector to provide a relationship between the state vector and the noise free measurement.

(4) The covariance matrices of all the signatures are averaged. This is identified as an average estimate of the measurement noise covariance, R, as required for the Kalman filter.

(5) An augmented matrix is formed by replicating and scaling the matrix R. This augmented matrix is identified as the covariance Q of the successive differences in the state vector. Covariance Q is assumed to be a simple function of R, and this assumption results in significant savings in computation time, since matrix inversions are not required for each update, and the computer memory requirements are minimal.

In the remainder of this report, we shall describe in some detail our approach, starting with the basic formulation of a Kalman filter. The basic formulation is then expanded to increase its versatility and utility. We shall describe the adapting procedures used previously and compare these procedures with the present formulation. We shall also show some test results from two test sets which demonstrate some of the capabilities of the new procedure. The appendices describe the mathematical background to the Kalman filter. A separate section is included to present the theory and some test results obtained by using subsets of linear combinations of channels. Limited test results indicate that the use of a subset of linear combinations may be a feasible method of reducing processing costs without any sacrifice in average classification accuracy.

3
USE OF A DECISION-DIRECTED KALMAN FILTER FOR ADAPTIVE PROCESSING

In this section, we shall show a general formulation of a Kalman filter and discuss in qualitative terms the way such a filter may be used to update the signature mean vectors. Various terms of the equations can be identified in ways we will demonstrate with observed phenomena. Next, we shall simplify the equations into a basic form and show explicitly how one would use them to update the mean vectors; this updating depends upon certain statistics of the data. One possible approximation to the statistics leads to a simplification of the equations along with greatly reduced computational and memory requirements of a general purpose digital computer. Finally, we shall show some extensions of the simplified filter that appear at this time to be most useful. In the discussion of the Kalman filter that follows, we shall isolate the filter from the remaining processing operations. When we say that a quantity (e.g., a decision) is known, we mean that that quantity is obtained from another portion of the processor or possibly from separate measurements.

3.1 BASIC FILTER EQUATIONS

We shall begin our discussion with one version of the Kalman filter equations. Consider the linear vector difference equation

$$x_k = \Phi_{k,k-1}x_{k-1} + \Gamma_{k,k-1}u_{k-1} + f_{k-1} + w_{k-1} \tag{1}$$

The purpose of the Kalman filter is to estimate the state vector x_k. The matrix $\Phi_{k,k-1}$ is known as a transition matrix and has the properties

$$\Phi_{k,k} = I \tag{2}$$

$$\Phi_{kj}\Phi_{ji} = \Phi_{ki} \tag{3}$$

$$\Phi_{k,j}^{-1} = \Phi_{jk} \tag{4}$$

$\Phi_{k,k-1}$ can be used to express known multiplicative changes in the state vector. The vector u_k is known as a control vector and is related to the state vector by the matrix $\Gamma_{k,k-1}$. The vector f_k is a forcing function, which, together with $\Phi_{k,k-1}$, can be used to introduce known additive and multiplicative changes to the state vector. The vector w_k is a sample from a random process with statistics:

$$E(w_k) = 0 \tag{5}$$

$$E\left(w_k w_j^t\right) = Q_k \delta_{k,j} \tag{6}$$

where E( ) denotes the expected value, and Q_k is the covariance matrix of the state variation.

In addition to the state equation, we have a measurement equation:

$$z_k = H_k x_k + v_k \tag{7}$$

where z_k is the measurement vector, which we can identify with the data

H_k is a known observation matrix

v_k is measurement noise with the properties:

$$E(v_k) = 0 \tag{8}$$

$$E\left(v_k v_j^t\right) = R_k \delta_{k,j} \tag{9}$$

$$E\left(v_k w_j^t\right) = 0 \tag{10}$$

where R_k is the covariance matrix of the measurement noise. Equations (1) and (7) define a general formulation of an estimation problem which we could use. We shall identify the components of the state vector x_k with the means of all of the materials. The random vector w_k represents inter- and intra-field variations that have been observed in the data, as well as gradual changes caused by phenomena such as atmospheric, sun position, and scanner variations. It is in the estimation of Q_k, the covariance matrix of w_k, that we were able to simplify the processing requirements.

Let us now formulate the complete Kalman filter problem, with a simplified form of Eq. (1)

$$x_k = x_{k-1} + w_{k-1} \tag{11}$$

The recursive estimate $\hat{x}_k$ of the state vector x_k is

$$\hat{x}_k = \hat{x}_{k-1} + K_k(z_k - H_k \hat{x}_{k-1}) \tag{12}$$

The matrix K_k is known as the Kalman filter and minimizes $E\left(\tilde{x}_k^t \tilde{x}_k\right)$, where

$$\tilde{x}_k = \hat{x}_k - x_k \tag{13}$$

is the difference between the true state vector x_k and our estimate $\hat{x}_k$. The symbol E denotes the expectation operator. We can expect K_k to depend on H_k and the covariance matrices Q_k and R_k. It is shown in Appendix A that

$$K_k = P_k' H_k^t \left[H_k P_k H_k^t + R_k\right]^{-1} \tag{14}$$

where

$$P_k = P'_k - K_k H_k P'_k \tag{15}$$

$$P'_k = P_{k-1} + Q_{k-1} \tag{16}$$

The recursion relationship is defined by Eqs. (12, 14, 15 and 16). When a new measurement z_k is received, the computer has P'_k and $\hat{x}_{k-1}$ in memory from the previous calculation. We also know the observation matrix H_k. We have all of the information to compute $\hat{x}_k$, P_k and P'_{k+1} and are ready to repeat the process for z_{k+1}.

Two features of this recursion relationship should be clarified. We assume that we know H_k, which means that we know the particular material or class that was present when z_k was measured. But this is not the case, because all we have present is the decision [15] or estimated mixture [16-18] for that measurement. We shall assume that we have the decision available, and shall use the decision, as though it were correct, to formulate z_k. This use of the decision in the Kalman filter has been called decision-directed.

The other feature to clarify is the first step of the recursive relationship. When the training data is at the beginning of the data to be processed, we can let P_0 be the zero matrix and $\hat{x}_0$ be the means of the signatures. When the training data is not near the beginning of the data, we must increase P_k to reflect the variations in the mean that could occur prior to processing of the training data.

One of the modifications to the basic Kalman filter equations which we shall describe in Section 3.9 and Appendix B shows a way to look ahead so that the information from training fields downstream can be used for the k-th observation.

We next consider the covariance matrices to be used. For the matrix R_k associated with the measurement noise v_k, we shall use the average covariance matrix of all of the signatures. If each covariance matrix for each of the classes is R_i, $i = 1, \ldots, m$, then we shall use

$$R_k = \frac{1}{m}\sum_{i=1}^{m} R_i = R \tag{17}$$

Two assumptions are inherent in Eq. 17: (1) the covariance matrices are sufficiently similar that one common matrix can be used for all classes, and (2) that all of the variation measured in the training data can be attributed to measurement noise rather than state variations. The use of the first assumption is justified when we find an improvement in recognition performance. The second assumption will be shown not to be restrictive because the important relationship is the measurement covariance matrix R_k relative to the state covariance matrix Q_k.

Q_k is the covariance matrix of w_k, the state variation. Because the state vector x_k is composed of the signature means, Q_k describes the slow variations of these means. We must

describe how the means of each material can vary and also how the variation in the means of different channels are correlated. We shall assume that Q_k can be written in the form

$$Q_k = \theta \otimes R = Q \tag{18}$$

where the operation $\otimes$ is the Kronecker product

$$\theta \otimes R = \begin{pmatrix} \theta_{11}R & \theta_{12}R \ldots \\ \theta_{21}R & \\ \cdot & \\ \cdot & \\ \cdot & \end{pmatrix} \tag{19}$$

of partitioned matrices. As we shall see, our assumption of the form of Q_k will lead to a simplification in the processing.

3.2 REDUCED EQUATIONS

Let us now see how this simplification arises. We first note that for our decision-directed filter, we can write H_k as

$$H_k = M_k^t \otimes I_n \tag{20}$$

where I_n is the identity matrix with rank n, the number of channels, and M_k is a column vector with a one indicating a decision and m-1 zeros. For example, if the fifth data point were classified as the third of six possible classes, then

$$M_5 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{21}$$

The following theorem is needed for the simplification. Let

$$P_1' = \phi_1 \otimes R, \quad H_k = M_k^t \otimes I_n, \quad Q_k = \theta \otimes R \tag{22}$$

Then

$$P_k' = \phi_k \otimes R \tag{23}$$

Where

$$\phi_{k+1} = \phi_k - \frac{\phi_{km}\phi_{km}^t}{\phi_{kmm+1}} + \theta \tag{24}$$

$$\phi_{km} = \phi_k M_k \tag{25}$$

$$\phi_{kmm} = M_k^t \phi_k M_k \tag{26}$$

We shall prove this theorem by induction. We first note that it is true for k = 1. We assume the theorem true for k = α. Then, from Eq. (14)

$$K_\alpha = (\phi_\alpha \otimes R)(M_\alpha \otimes I_n)\left[\left(M_k^t \otimes I_n\right)(\phi_\alpha \otimes R)(M_\alpha \otimes I_n) + R\right]^{-1}$$

$$= (\phi_{\alpha M} \otimes R)[(\phi_{\alpha MM} + 1) \otimes R]^{-1} = \frac{\phi_{\alpha M}}{\phi_{\alpha MM+1}} \otimes I_n \tag{27}$$

When we evaluate Eq. (15), we have

$$P_\alpha = \phi_\alpha \otimes R - \left(\frac{\phi_{\alpha M}}{\phi_{\alpha MM+1}} \otimes I_n\right)\left(M^t \otimes I_n\right)(\phi_\alpha \otimes R) = \left(\phi_\alpha - \frac{\phi_{\alpha M}\phi_{\alpha M}^t}{\phi_{\alpha MM+1}}\right) \otimes R \tag{28}$$

From Eq. (16) we have

$$P'_{\alpha+1} = P_\alpha + Q_\alpha = \left(\phi_\alpha - \frac{\phi_{\alpha M}\phi_{\alpha M}^t}{\phi_{\alpha MM+1}} + \theta\right) \otimes R \tag{29}$$

Q.E.D. We can now evaluate Eq. (12).

$$\hat{x}_k = \hat{x}_{k-1} + \left(\frac{\phi_{km}}{\phi_{kmm+1}} \otimes I_n\right)(z_k - y_k) \tag{30}$$

where $y_k = H_k \hat{x}_{k-1}$ (31)

is the estimate of the mean of the material that was recognized.

The updating equations are now contained in Eqs. (24) and (30). We need to store an mn×1 state vector $\hat{x}_k$, which is already stored as part of the signatures, and an m×m symmetric matrix ϕ_k. We also need θ, which we shall assume has the form

$$\theta = \theta_1 \begin{pmatrix} 1 & \theta_2 & \cdots \\ \theta_2 & 1 & \\ \cdot & & \\ \cdot & & \\ \cdot & & \end{pmatrix} \tag{32}$$

Only two scalars are needed to determine θ. Because the updating equations are now independent of R, we can determine the operation of the filter by choice of θ_1 and θ_2. We shall choose

a value of θ_2 in the range $0 \leq \theta_2 < 1$, because θ_2 determines the amount of correlation of the variations in the signature means that we have assumed. We shall now develop a simple explanation for θ_1.

3.3 UPDATING RATE

Let us assume that we always recognize the first class. Let us also assume that the measurements z_k are all zero vectors for a sufficient number of samples so that our estimate of the first mean vector is also the zero vector. We now let the measurements z_k become a vector composed of ones, and compute the number of updates needed for the estimate of the mean to equal $1 - e^{-1}$. We shall call this number of updates the updating rate and show that it is simply related to θ_1.

We need to consider only the first element of $\hat{x}_k$ and the elements in the first row and column of θ and ϕ_k. From Eq. (24), the ϕ_{11} element of ϕ_k approaches a constant value when $\phi_{k+1} = \phi_k$, or

$$\frac{\phi_{11}^2}{\phi_{11} + 1} = \theta_1 \tag{33}$$

$$\phi_{11} = \frac{\theta_1 + \sqrt{\theta_1^2 + 4\theta_1}}{2} \tag{34}$$

The updating equation for the first element of x_k is

$$\hat{x}_{k,1,1} = \hat{x}_{k-1,1,1} + a(z_o - \hat{x}_{k-1,1,1}) \tag{35}$$

where

$$a = \frac{\phi_{11}}{\phi_{11} + 1} \tag{36}$$

We let $\hat{x}_o = 0$, and $z_k = 1$, $k = 1, \ldots$, then, from Eq. (35), we have

$$\hat{x}_{k11} = [1 - (1 - a)^k] = 1 - \left[1 + \frac{\theta_1}{2} - \frac{\sqrt{\theta_1^2 + 4\theta_1}}{2}\right]^k \tag{37}$$

For the values of θ that we use ($\theta_1 \leq 10^{-3}$), we can approximate Eq. (37) with

$$\hat{x}_{k11} \approx 1 - (1 - \sqrt{\theta})^k \tag{38}$$

Thus, when

$$k = \frac{1}{\sqrt{\theta}} \tag{39}$$

we have

$$\hat{x}_{k11} \approx 1 - e^{-1} \tag{40}$$

For example, when $\theta_1 = 10^{-3}$ and $k = \frac{1}{\sqrt{\theta_1}}$, the difference between the results obtained by using Eqs. (37) and (40) is 1.5×10^{-5}. We conclude that Eq. (39) is sufficiently accurate for our purposes.

3.4 OTHER FORMS OF UPDATING

Let us look briefly at the previously used adaptive procedures and see how they can be interpreted in terms of our Kalman filter formulation [7, 8]. Two methods of adapting the signature means were examined: non-interactive and interactive. The non-interactive updating can be implemented by setting

$$\phi_k = \frac{1}{W} I_m \tag{41}$$

The scalar W was used to vary the updating rate. Note that ϕ_k is independent of k. This implies that there is variation only in the mean of the material that was recognized, and all other means did not change. For approximately equal numbers of decisions for each material, this method should produce estimates of the mean vectors that are similar to those obtained by the basic Kalman filter.

For the interactive method, we would set

$$\phi_k = \frac{1}{W-1} \begin{pmatrix} 1 & \frac{\rho_2}{\rho_1} & \frac{\rho_3}{\rho_1} & \cdots \\ \frac{\rho_1}{\rho_2} & 1 & \frac{\rho_3}{\rho_2} & \\ \vdots & & & \end{pmatrix} \tag{42}$$

where ρ_i, $i = 1, \ldots, m$ is a scalar constant for each material. Again, ϕ_k is independent of k. In addition, ϕ_k is not symmetrical, implying a lack of symmetry in the covariance matrix ρ_k. Thus, we must conclude that either the interactive updating is non-optimal or the model described by Eqs. (1) and (7) is not the only reasonable model for updating. However, the interactive method was developed specifically because random multiplicative, rather than additive, changes in the mean vectors were assumed to be present. We should note that the interactive method is easily expressed in the Kalman filter formulation.

3.5 LINE-BY-LINE UPDATING

We have now finished our discussion of the simplified Kalman filter. We have developed a filter that can be used to update the mean vectors each time a data point is classified. As we shall see when we discuss test results, a relatively slow updating rate will produce the lowest classification error rate. Knowing this, we are led to consider the possibility of updating after each line or portion of a line, rather than after each data point. Our motivation is to reduce processing time, even though the updating requires fewer calculations than does the decision rule. One might hope to save additional computer time by using a linear [Refs. 20, 21] rather than a quadratic (maximum likelihood) decision rule. The problem with using a linear decision rule with a Kalman filter is that some computation for each point using mean vectors and covariance matrices is required before one is ready to apply the rule. This computation is only required once for non-adaptive processing, and therefore little increases the overall processing time in the non-adaptive linear decision rule case. But it appears this computation is time consuming in the Kalman filter point-by-point updating case so that little if any time is saved.

3.6 PREPROCESSING

We can add another function to our filter. We have not considered variations in the data that are dependent on the angle of view (scan angle). One would like to preprocess the data to remove or reduce any scan angle dependence before making classification decisions. It is not unlikely that whenever the mean vectors change, the scan angle dependence and the required preprocessing would also change. Therefore, we have a motive for using the Kalman filter to update the scan angle correction function as it updates the mean vectors.

Many forms of preprocessing have been studied. One is particularly suited for inclusion in the Kalman filter equations. One can estimate the parameters of a polynomial function of scan angle (a separate polynomial is needed for each data channel) that would be added to the data so that the resultant sum would be approximately independent of angle. Another use of the polynomial would be to subtract it from the mean vectors. Either use produces the same decision. One can either use one polynomial to be applied to all classes or separate polynomials, one for each class. The latter method is preferred when we expect the various classes to have different angular dependences. A disadvantage of the latter method is that we must use a reduced updating rate, because there are more parameters to estimate (more polynomials) and when we use multiple parameters to describe any mean vector, there is a reduced maximum error that can be tolerated in our estimate of each of the parameters before classification accuracy is degraded rather than improved.

3.7 CONFIDENCE FACTOR

In addition to updating every portion of a line of data and estimating a scan angle correction term, we shall make use of a confidence factor. A confidence factor has already been used successfully [Ref. 12]. The purpose of the confidence factor is to weight the contribution that each data point makes to the updating by a number that represents our confidence that we have made the correct decision. A point very close to a mean vector is weighted more than a more distant point. We would be trying to reduce the effect of using a decision-directed filter which occasionally used the incorrect decision. We have no model to use to determine the weights, but chose a method that is computationally simple. Because we always compute the χ^2 value for each data point after a decision is made or during the decision process, we can use as a weight the probability (from the χ^2 distribution) that a data point from a Gaussian distribution would have the calculated value or larger. Thus we are using the complement of the χ^2 accumulative distribution function.

3.8 REVISED UPDATING EQUATIONS

We shall now introduce a formulation that we chose to augment our first improvements to the basic Kalman filter. We shall model our system by Eq. (11), although we now have the following form for x_k

$$x_k = \begin{pmatrix} \mu_{1k} \\ \cdot \\ \cdot \\ \cdot \\ \mu_{mk} \\ P_{11} \\ \cdot \\ \cdot \\ \cdot \\ P_{Vk} \end{pmatrix} \tag{43}$$

where μ_{ik} is the mean vector of the i-th material, and P_{ik} is the i-th vector (one element for each channel) of the polynomial, when the k-th data point was measured. We replace Eq. (7) with

$$z_{ki} = H_{ki} x_k + v_{ki} \tag{44}$$

where the subscript i refers to the measurement at the i-th scan angle α_i. We shall use α_i to determine H_{ki}:

$$H_{ki} = \left(\mu_{ki}^t \; \alpha_i \alpha_i^2 \ldots \alpha_i^V\right) \otimes I_n \tag{45}$$

To better understand Eq. (45), let us evaluate Eq. (44) if the j-th material were chosen. We would have

$$Z_{ki} = \mu_{jk} + \alpha_i P_{1k} + \alpha_i^2 P_{2k} + \ldots + \alpha_i^V P_{Vk} + v_{kj} \tag{46}$$

which means that we assume that we are measuring a random vector $(\mu_{jk} + v_{kj})$ added to a polynomial function of angle that is independent of the material. So that we shall be consistent with Eqs. (8), (9) and (10), we get

$$E(v_{ki}) = 0 \tag{47}$$

$$E\left(v_{ki} v_{1j}^t\right) = \delta_{k1}\delta_{ij}R_k \tag{48}$$

$$E\left(v_{Rj} w_1^t\right) = 0 \tag{49}$$

We now rewrite Eq. (12)

$$\hat{X}_k = \hat{X}_{k-1} + K_k \sum_{i=1}^{N} C_{ki}(Z_{ki} - H_{ki}\hat{X}_{k-1}) \tag{50}$$

We have included the confidence factor C_{ki} for each data point z_{ki} and show our intention of updating after every N data point. The Kalman filter K_k will differ from that used previously. We also have to use a new interpretation of w_k, because it now represents a change that could occur after N data points rather than after each data point. We shall now find the K_k that we should be using. Our method will be to show that we can rewrite the new equations into the same equations that we had previously, and then to use the same solution. Our new filter will have the same form as the previous one, although the terms will be interpreted differently. Actually, our method of solution sounds more complicated than it really is. We define

$$H_k = \sum_{i=1}^{N} C_{ki} H_{ki} \tag{51}$$

$$z_k = \sum_{i=1}^{N} C_{ki} z_{ki} \tag{52}$$

$$v_k = \sum_{i=1}^{N} C_{ki} v_{ki} \tag{53}$$

$$M_k^t = \sum_{i=1}^{N} C_{ki} M_{ki}^t \sum_{i=1}^{N} C_{ki}\alpha_i \cdots \sum_{i=1}^{N} C_{ki}\alpha_i^V \tag{54}$$

From Eqs. (51) and (54), we have

$$H_i = M_k^t \otimes I_n \tag{55}$$

We are now in the position to rewrite Eqs. (44) and (50).

$$z_k = H_k x_k + v_k \tag{56}$$

$$\hat{x}_k = \hat{x}_{k-1} + K_k(z_k - H_k\hat{x}_{k-1}) \tag{57}$$

We now use Eq. (48) to compute the statistics of our new v_k.

$$E(v_k) = 0 \tag{58}$$

$$E\left(v_k v_j^t\right) = \delta_{kj} R_k \sum_{i=1}^{N} C_{ki}^2 \tag{59}$$

Thus our updating equations become

$$\hat{x}_k = \hat{x}_{k-1} + \frac{\phi_{km}}{\phi_{kmm} + \sum_{i=1}^{N} C_{ki}^2} \otimes I_n \, (z_k - H_k\hat{x}_{k-1}) \tag{60}$$

$$\phi_{k+1} = \phi_k - \frac{\phi_{km}\phi_{km}^t}{\phi_{kmm} + \sum_{i=1}^{N} C_{ki}^2} + \theta_k \tag{61}$$

where, as before,

$$\phi_{km} = \phi_k M_k \tag{62}$$

$$\phi_{kmm} = M_k^t \phi_{km} \tag{63}$$

Note that we use Eqs. (51), (52), and (54) to find $H_k z_k$ and M_k. When we classify the data, we replace each mean vector in the signatures with the sum of the mean vector and the polynomial vector.

We have one remaining step, that of replacing Eq. (32) with a new estimate of θ_k. Our choice is

$$\theta_k = N\theta_1 \left(\begin{array}{ccc|c} 1 & \theta_2 & \cdots & \\ \theta_2 & 1 & & 0 \\ \cdot & & & \\ \cdot & & & \\ \hline 0 & & & \gamma_1 I_V \end{array}\right) \tag{64}$$

We now have three parameters, θ_1, θ_2, and γ_1 that we can use to determine θ_k. We chose to use $N\theta_1$ as a multiplier so that we could interpret θ_1 in the same way we did previously. We are also assuming that the mean vectors and the parameters of the polynomial are uncorrelated, as well as the vectors of the different orders of the polynomial.

We could determine the operation of the modified Kalman filter by the three parameters (θ_1, θ_2, and γ_1) indicated in Eq. (64). We choose to add a fourth parameter, a confidence scale factor, which we used as a divisor of the X^2 values for the data points before determining the C_{ki} confidence factors. With this fourth parameter we can alter the rapidity with which the confidence factor decreases as the data points become more distant from the mean vectors.

There is one problem when implementing the reduced updating equations. When the scan angle correction terms are added to the state matrix, we must be sure to have sufficient independent measurements so that we can estimate each term of the state vector. Without the scan angle terms, this independence requirement is met when the H_k vary (as they normally do). With the scan angle terms, however, extra precautions are needed. We have chosen to update V + 1 times per line for a V-th order polynomial.

3.9 USE OF AUXILIARY GROUND TRUTH

Before concluding this section, we shall consider a method of reducing an undesirable characteristic of any updating method, including the Kalman filter, known as signature capture. In multispectral scanner data processing, signature capture occurs when the mean of one class actually describes the data from a different class. We would tend to recognize data from one (or more) material incorrectly, clearly an undesirable situation. Of course, the same misclassification can occur without updating, and it is possible that the updating may eliminate the problem. As we shall see, we tend to decrease the probability of misclassification by increasing the updating rate up to a point. There appears to be a maximum rate beyond which the probability of misclassification increases. The increase appears to be caused by capture. We believe that we reduce the probabilities of both misclassification and capture when we use the confidence factor. Another method of reducing the probability of capture is now described.

One method of processing data does not use updating. Training fields are located near the beginning of the data collection pass to be processed, and auxiliary training fields are identified throughout the run, so that the signatures can be changed occasionally. We could use a variant of the same technique. Hopefully, fewer auxiliary training fields would be needed, because the updating fulfills the task of adapting the signatures to the changing statistics of the data. The main purpose of the additional ground truth would be to reduce the probability of the occurrence of a capture.

It should be possible to reformulate the Kalman filter equations to use the additional ground truth statistics. However, we lose nothing by finding the optimum filter for a separate formulation. The derivation of the filter equations for this case can be found in Appendix B. A brief description of the change in the updating characteristics is in order here.

The most obvious change is that the complexity of the filtering equation is increased. We have not tested this particular modification to the filter, so we are not in a position to evaluate the extent of the increase in complexity. Another change is that we have formulated the problem so that the updated mean of a class is exactly the mean of the training field for that class when the center of the field is being processed. The influence of a training field decreases as the distance from the training field to the ground represented by the data being processed increases. We could have assumed some uncertainty in our knowledge of the mean vectors of the training fields, but we saw no advantage in doing so.

One requirement that we have placed on the operation of all of our formulations of the Kalman filter is that the training fields be located near the beginning of the run. We may be able to relax this requirement when we use the initial training fields themselves as auxiliary training fields.

4
EXPERIMENTAL EVALUATION OF SOME KALMAN FILTER CONFIGURATIONS

The purpose of this section is threefold: (1) to show the results of some tests that were performed to validate the theory presented in the last section, (2) to describe problem areas that are not in our analytical model, and (3) to list the values of the parameters, e.g., θ_1, θ_2, that should be used in the Kalman filter.

An important consideration that was faced was that of choosing data sets. Previously, testing of adaptive processing at ERIM was limited to aircraft data over the Indiana C-3 area gathered in 1966 with the M-5 scanner [11, 12]. This data was originally chosen because it had been noted that classification accuracy tended to be significantly better near the beginning than near the end of the flightline. It was hoped that this loss of accuracy, which was observed in a short run, would be representative of a loss that might occur in processing a long data run.

There were two disadvantages to the use of this data: (1) all of our test had been restricted to this one data set, and (2) the data were collected with an outdated multispectral scanner. There was one advantage, however, in the fact that the previous tests had been made using this data. We would be able to compare the Kalman filter results with the previous results. This comparison would be useful in ensuring that the Kalman filter was operating correctly.

4.1 C-3 TEST RESULTS

We chose to perform our preliminary tests on the C-3 data, and to use a different data set to substantiate our findings. We believe that this choice was a good one, even though we were unable to make a decisive comparison between the operation of the Kalman filter and the techniques used previously. Previous evaluations had been made by manual analysis and comparisons of digital recognition maps. This method was adequate for its purpose, that of demonstrating the usefulness of adaptive processing in a clear-cut case. The method lacked the ability to differentiate between small differences in performance.

Our first task was to devise a method of comparing different decision techniques that would improve sensitivity to small differences. We felt the need to display the recognition results so that we could readily compare a given adaptive technique with any other, and could find any peculiarities that might be present. One method that we tried is demonstrated in Fig. 1, where we have shown the classification accuracies that were measured for each field relative to the along-track position of the field. To obtain our figures we had to identify portions of the data with specific fields, using only fields for which ground truth was available. We were able to eliminate from the presentation classification results for fields for which no ground truth or signatures were available, as well as field boundaries.

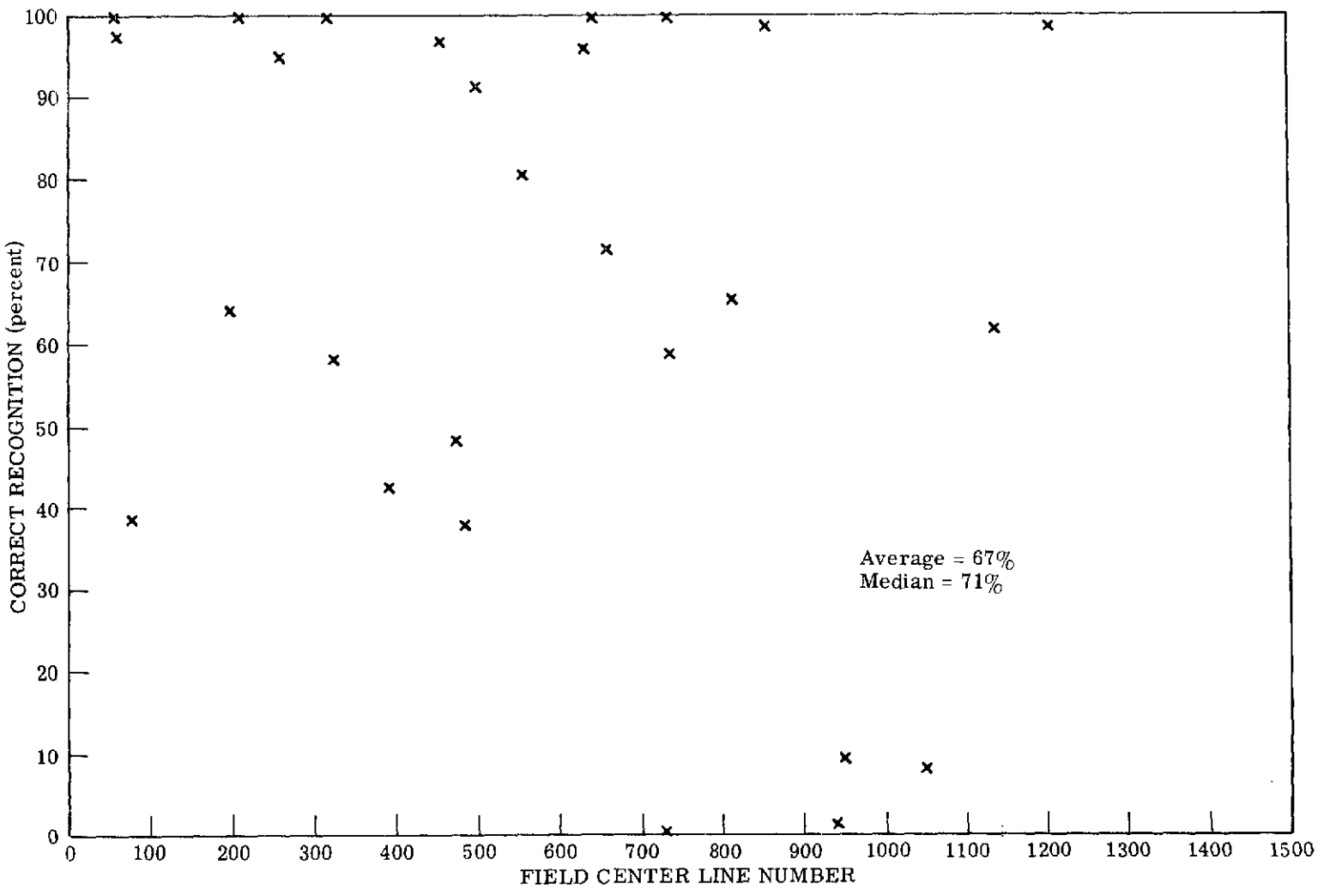


FIGURE 1. CLASSIFICATION RESULTS WITHOUT UPDATING

Figure 1 shows the classification results that were obtained when no adaptive processing was employed. There are four fields for which the classification accuracy was less than 10%. There was no capture, obviously, because there was no adapting. We believe that the data from those four fields did not follow a general processing assumption, that the data from each class would have statistics similar to those obtained from the training data. It appears that it is rather difficult to define capture quantitatively, because capture is normally considered to be a result of adapting.

We have presented two additional classification results in Figs. 2 and 3. Both figures were obtained using a basic Kalman filter, updating after each data point. The difference between the two figures is the result of using different values for the parameters θ_1 and θ_2. We found that we had difficulty in choosing which set of parameters would be preferable. In fact, we could not always convince other investigators that either one of the updated results was preferable to those obtained without updating.

Clearly, another method was needed. We computed the mean and median values for the recognition accuracies shown in Figs. 1-3. In finding the median, we noted that the value we obtained depended to some degree on the recognition accuracy of one field. The mean, on the other hand, did not exhibit this phenomenon, and probably better represents some user requirements. Consequently, we chose the mean classification accuracy as a measure of performance.

A total of 27 fields were identified for our tests from the C-3 data. Four of the fields, one each for corn, wheat, pasture, and soybeans, are used for training. Initial tests indicated a sudden change in the characteristics of the data after the first 14 fields. Consequently, we have subdivided the 27 fields into three groups: (1) the four training fields, (2) the ten non-training fields in the first 14, and (3) the remaining 13 fields. There are materials represented in the data other than the four materials chosen for testing. This data, when improperly classified as one of the four test materials, provides false input to the Kalman filter.

For our initial tests, we used the Kalman filter that updated after each data point and did not use a confidence factor. Our first tests were made with different values of θ_1 (which determines the updating rate) and θ_2 (the interaction of updating the means of different classes). In Table 1 we show the classification accuracies obtained for different values of θ_2 with θ_1 constant. We include the accuracies obtained without updating for comparison. The results show that the accuracies are not very dependent on our choice of θ_2. Consequently, we decided that we should be using $\theta_2 = 0.7$. Note that the accuracies are higher when updating is used.

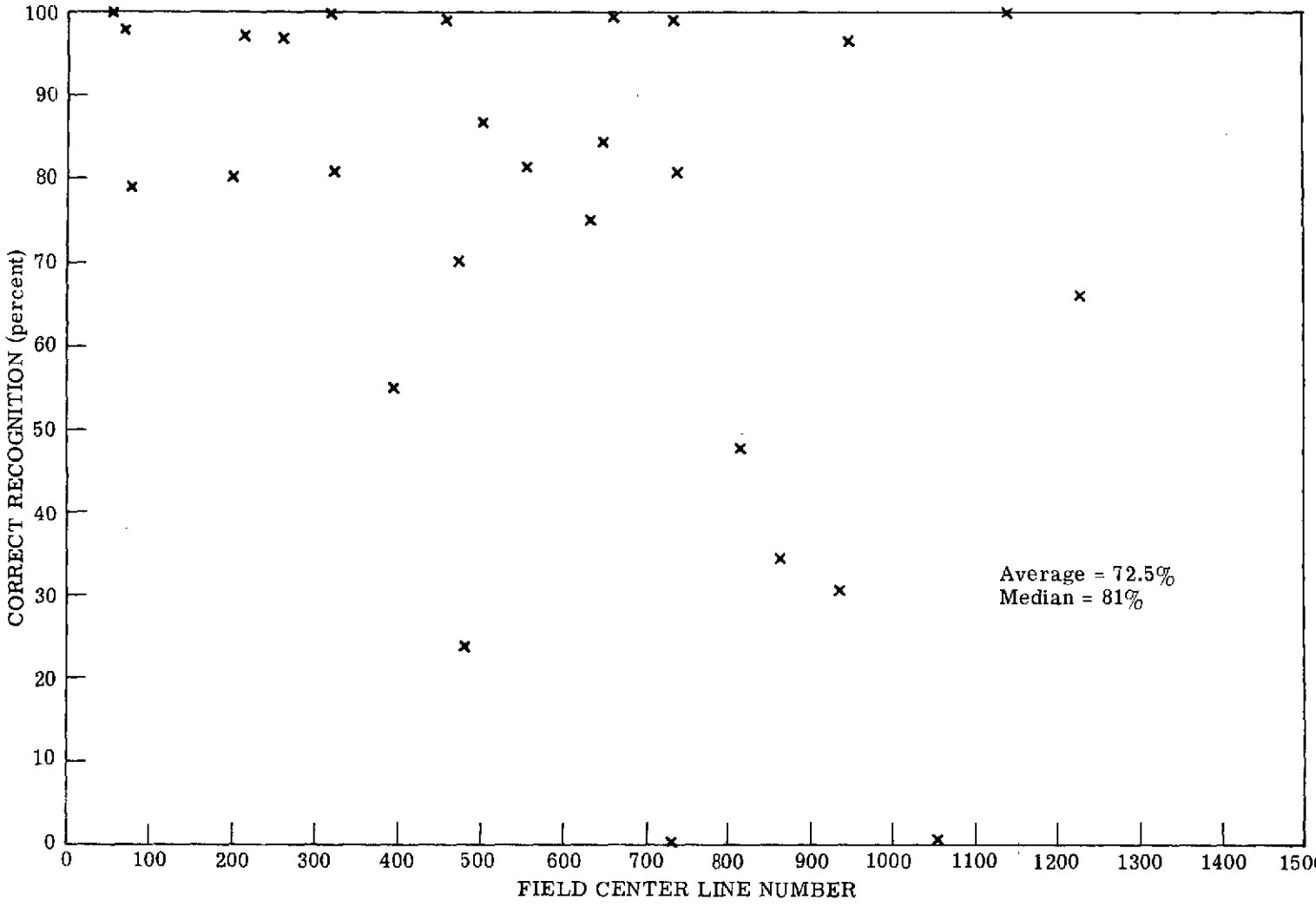


FIGURE 2. CLASSIFICATION RESULTS WITH $\theta_1 = 10^{-6}$, $\theta_2 = 0.5$

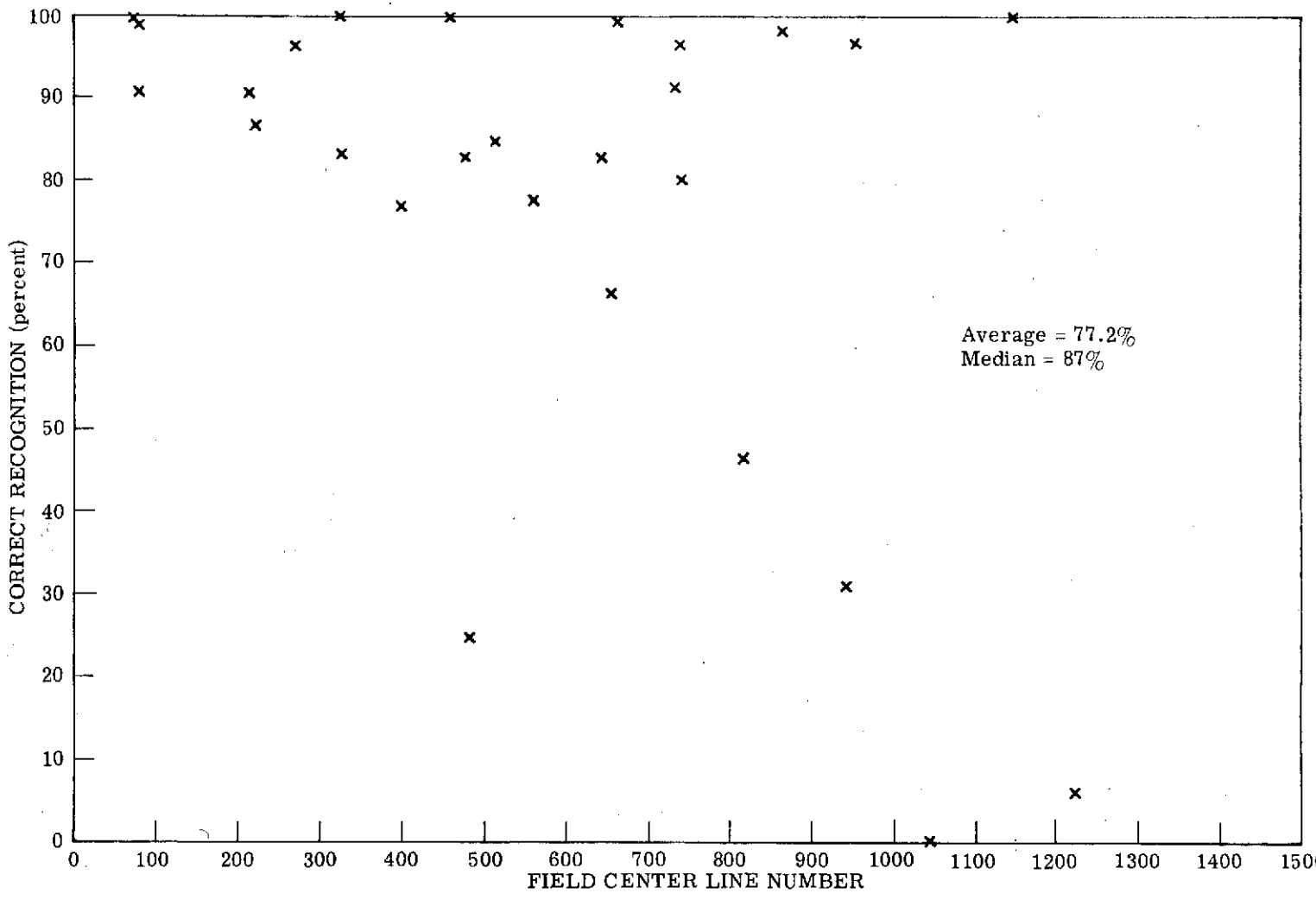


FIGURE 3. CLASSIFICATION RESULTS WITH $\theta_1 = 10^{-5}$, $\theta_2 = 0.7$

The effects of using different values of θ_1 are shown in Table 2. Again we include the accuracies that were obtained when no updating was used. Improved accuracies resulted from all values of θ_1 except $\theta_1 = 10^{-4}$. It appears that 10^{-4} is too large a value for θ_1 (the updating rate is too large). The loss of accuracy may be caused by capture, which can occur most easily when θ_1 is large.

There appear to be two trends in the data, one for the last 13 test fields, where a small value of θ_1 (10^{-7}) is preferable, and the other for the first two categories, where a larger value of θ_1 (10^{-5}) is preferable. We have one explanation for these trends. When we test the first two categories, we may be adapting the means to the local inter- and intra-field variations. On the other hand, the accuracies obtained from the last 13 test fields may indicate that for gradually changing data, it is better to update slowly and not try to follow rapid data changes. If this explanation is correct, it might mean that one should use a combined rapidly and slowly changing filter characteristic, which would be accomplished by using a Kalman filter designed for non-white state variations.

A comparison was made between the accuracies obtained when updating every line and every point. The results are shown in Table 3. We do not believe that the differences in accuracies are significant, and have concluded that updating every line is the preferred method because of the reduced processing time.

Our final tests on the C-3 data set were made to evaluate the confidence factor, updating every line. The results are shown in Tables 4 and 5. In Table 4, we show the results for the first 10 test fields, for different values of θ_1 and the confidence scale factor. The values shown for a confidence scale factor of infinity are repeated from Table 2, where the confidence factor was held constant at unity (no confidence factor). There appears to be no improvement to be gained by using the confidence factor, with the exception of $\theta_1 = 10^{-4}$. This exception does not appear to be of interest, however, because greater improvement can be gained by letting $\theta_1 = 10^{-5}$ and not using the confidence factor.

In Table 5, we observe that we can gain accuracy by using the confidence factor. We are led to the conclusion that the confidence factor is useful in following slowly varying changes in the data, but not rapid variations. The choice of scale factor that one would use can be seen to depend on the value of θ_1. When the scale factor decreases, the updating is more dependent on data values closest to the means of the classes. Apparently this dependence has the effect of decreasing the probability of capture, so that larger values of θ_1 can be used.

4.2 EATON COUNTY TEST RESULTS

A second data set that we tested was gathered during an ERTS underflight at 5000 feet of Eaton County, Michigan on 25 August 1972. This data set was chosen because it was gathered

TABLE 1. EFFECT OF θ_2 ON RECOGNITION ACCURACY. $\theta_1 = 10^{-6}$

	θ_2 Values			
	0.3	0.5	0.7	No Update
Training data	93.3	93.5	93.6	90.5
First 10 test fields	77.2	77.3	77.4	68.9
Last 13 test fields	62.0	62.4	64.2	59.0

TABLE 2. EFFECT OF θ_1 ON RECOGNITION ACCURACY. $\theta_2 = 0.7$

	θ_1 Values				
	10^{-4}	10^{-5}	10^{-6}	10^{-7}	No Update
Training data	68.7	93.9	93.6	91.6	90.5
First 10 test fields	60.6	81.7	77.4	72.2	68.9
Last 13 test fields	11.8	61.8	64.2	71.6	59.0

TABLE 3. COMPARISON OF PERCENTAGE RECOGNITION ACCURACIES WHEN UPDATING EVERY POINT AND EVERY LINE FOR TEST FIELDS. $\theta_1 = 10^{-6}$ and $\theta_2 = 0.7$.

	Point	Line
First 10 test fields	77.4	77.6
Last 13 test fields	64.2	63.5

TABLE 4. AVERAGE PERCENTAGE CORRECT RECOGNITION FOR FIRST 10 TEST FIELDS UPDATED EVERY LINE USING CONFIDENCE FACTOR

Confidence Scale Factor	θ_1 Values 10^{-4}	10^{-5}	10^{-6}	10^{-7}
∞	60.6	81.7	77.4	72.2
4	—	63.5	66.7	68.7
1	—	66.5	69.0	69.4
0.4	66.7	69.5	69.5	69.1
0.25	68.2	78.1	69.3	69.0

TABLE 5. AVERAGE PERCENTAGE CORRECT RECOGNITION FOR LAST 13 TEST FIELDS UPDATED EVERY LINE USING CONFIDENCE FACTOR

Confidence Scale Factor	θ_1 Values 10^{-4}	10^{-5}	10^{-6}	10^{-7}
∞	11.8	61.8	64.2	71.6
4		72.0	73.4	70.5
1		72.0	77.0	63.1
0.4	72.6	74.5	61.7	59.7
0.25	72.3	63.9	60.3	59.2

over one of the longest (21 mile) continuous ground areas of any of our data sets and there was sufficient ground truth available. The first 11 miles had extensive ground truth, the remaining 10 miles less. The first portion of the data could be used to test both the ability of the Kalman filter to follow rapid variations in the data and the updating of the scan angle correction (preprocessing). All of the data could be used for testing for slower variations.

Tests were made to determine whether the results obtained with the C-3 data would be duplicated. The first consideration was the choice of θ_1 and θ_2 to provide highest classification accuracy. A comparison of the choices of θ_1 cannot be made directly, because θ_1 must be interpreted differently in the two data sets. We shall avoid confusion by calling the θ_1 variable for the second data set θ_1'.

When we tested the C-3 data set, we used every second point and every fourth line. Our updating rate becomes $1/\sqrt{\theta_1}$ points processed, which is equivalent to $8/\sqrt{\theta_1}$ data points or $8/N\sqrt{\theta_1}$ data lines (N data points/line). One method of comparison is to use points per update when a filter is designed for rapid changes in the data and to use lines per update otherwise.

We used every point when we tested the Eaton County data. The variance of each state variable was assumed to be $N\theta_1'$ even though we updated the state vector three times each line. The updating rate now becomes $1/3\sqrt{N\theta_1'}$ lines and $N/3\sqrt{N\theta_3'}$ points per update. One should compare θ_1' with $0.625\theta_1$ when considering rapid updating rates and with $0.25\theta_1$ otherwise. These factors were obtained by knowing that there were 226 and 360 points per line for the C-3 and Eaton County data, respectively.

We used 25 fields located throughout the first eleven miles of the area for our first series of tests. There were five classes: corn, trees, beans, wheat, and bare soil. Some results of the tests using data that had been preprocessed using one of our standard techniques (ACORN, Ref. 7) can be seen in Tables 6 and 7. We obtained Table 6 by using a confidence scale factor of 1 and different values of θ_1' and θ_2. The best choice of θ_2 is zero, a clear difference from the value of 0.7 that was best for the C-3 data. The value of zero for θ_2 is more consistent with the assumption that we are adapting for rapid changes in the data. The choice of 0.7 for θ_2 for classifying the C-3 data may be an artifact of the techniques used in 1966 to collect and store the data.

There is better agreement found when θ_1 and θ_1' are compared. The numbers we compare are $\theta_1' = 10^{-8}$ and $0.625\theta_1 = 0.625 \times 10^{-7} = 10^{-7.204}$. We did not find more precise values for optimum choices of θ_1', primarily because each number in the table represents approximately 36 minutes of IBM 7094 computer time, with the exception of the no update (maximum likelihood fixated signature) time, which took 11.3% less time.

TABLE 6. AVERAGE PERCENTAGE CORRECT RECOGNITION FOR 25 EATON COUNTY FIELDS USING A CONFIDENCE SCALE FACTOR OF 1.0

	θ_1' Values		
θ_2 Values	10^{-7}	10^{-8}	No Update
0	67.7	86.3	82.6
0.3	49.3	85.4	

TABLE 7. AVERAGE PERCENTAGE CORRECT RECOGNITION FOR 25 EATON COUNTY FIELDS USING A CONFIDENCE SCALE FACTOR OF 0.25

	θ_1' Values			
θ_2 Values	10^{-6}	10^{-7}	10^{-8}	No Update
0	84.5	85.5	83.9	
0.3	84.7	84.3	83.0	82.6
0.7	54.9	60.3	74.3	

Table 7 was obtained when we used a confidence scale factor of 0.25. Again the best choice that we found for θ_2 is zero. There is less agreement found after θ_1 and θ_1' are compared, because $0.625\theta_1 = 0.625 \times 10^{-5} = 10^{-5.204}$ and $\theta_1' = 10^{-7}$. However, there is no appreciable difference in results between $\theta_1' = 10^{-7}$ and $\theta_1' = 10^{-6}$. We tried using a confidence scale factor of 4 with $\theta_1 = 10^{-8}$ and $\theta_2 = 0$, but the average percentage correct recognition was so low (65.4%) that we did not continue using that confidence scale factor.

An additional test was made, which produced an interesting but unexpected result. For the test, we used our linear classifier, which does not have any adapting feature. We measured an average percentage correct recognition of 86.9%, which is higher than that measured when we used the quadratic decision rule with and without updating.

We do not feel that we have enough comparisons between the two decision rules to know whether this improvement can be expected in most data sets. It may be that when we use the linear decision rule with the Kalman filter, we may achieve even higher percentages, because the linear decision rule may be more robust. An equally reasonable explanation is that generally one cannot expect higher classification accuracy with a linear decision rule. However, we favor the first explanation.

We have included Figs. 4 and 5 to show examples of the recognition maps that one might obtain by using two different classification procedures. Figure 4 was obtained with our linear decision rule on preprocessed data. The average percentage correct recognition of 86.9% was obtained from 25 fields located within the map. Figure 5 was obtained by using the Kalman filter with $\theta_1' = 10^{-8}$, $\theta_2 = 0$, and a confidence scale factor of 0.25. From Table 6 we see that the average percentage correct recognition was 86.3%. Thus, the two measured accuracies are essentially the same. When one compares the two figures, one might conclude that fields tend to be more uniform in Fig. 5 (Kalman filter processing).

Our tests of the preprocessing feature of the new filter were inconclusive. The previous tests were made using preprocessed data. When we planned the tests, we discovered that two additional operational techniques would be needed. We needed some initial values for the polynomials (one for each channel) of the scan angle correction. The technique that we used was to use the filter to generate initial conditions by processing a portion of the data (200 lines), using initial values of zero for the polynomials, and using such a low value of θ_1 that there was negligible updating applied to the mean vectors. The estimates of the polynomial parameters were used as initial values for processing all of the data.

We also needed a method of obtaining signatures, including both first- and second-order statistics. The signatures that were used to obtain initial conditions were measured using the data before preprocessing was applied. We do not expect that these statistics would match

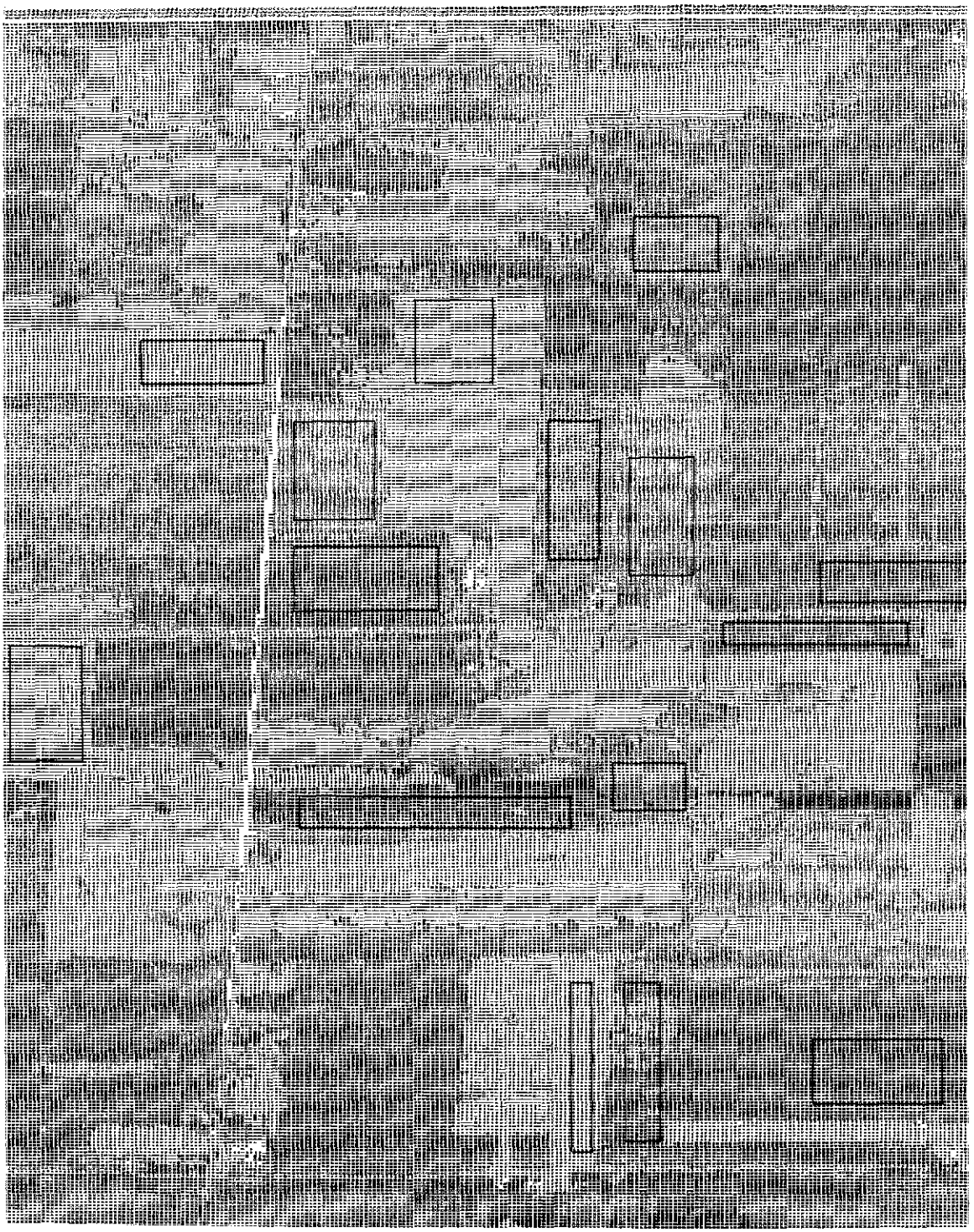


FIGURE 4. LINEAR DECISION RULE RECOGNITION MAP

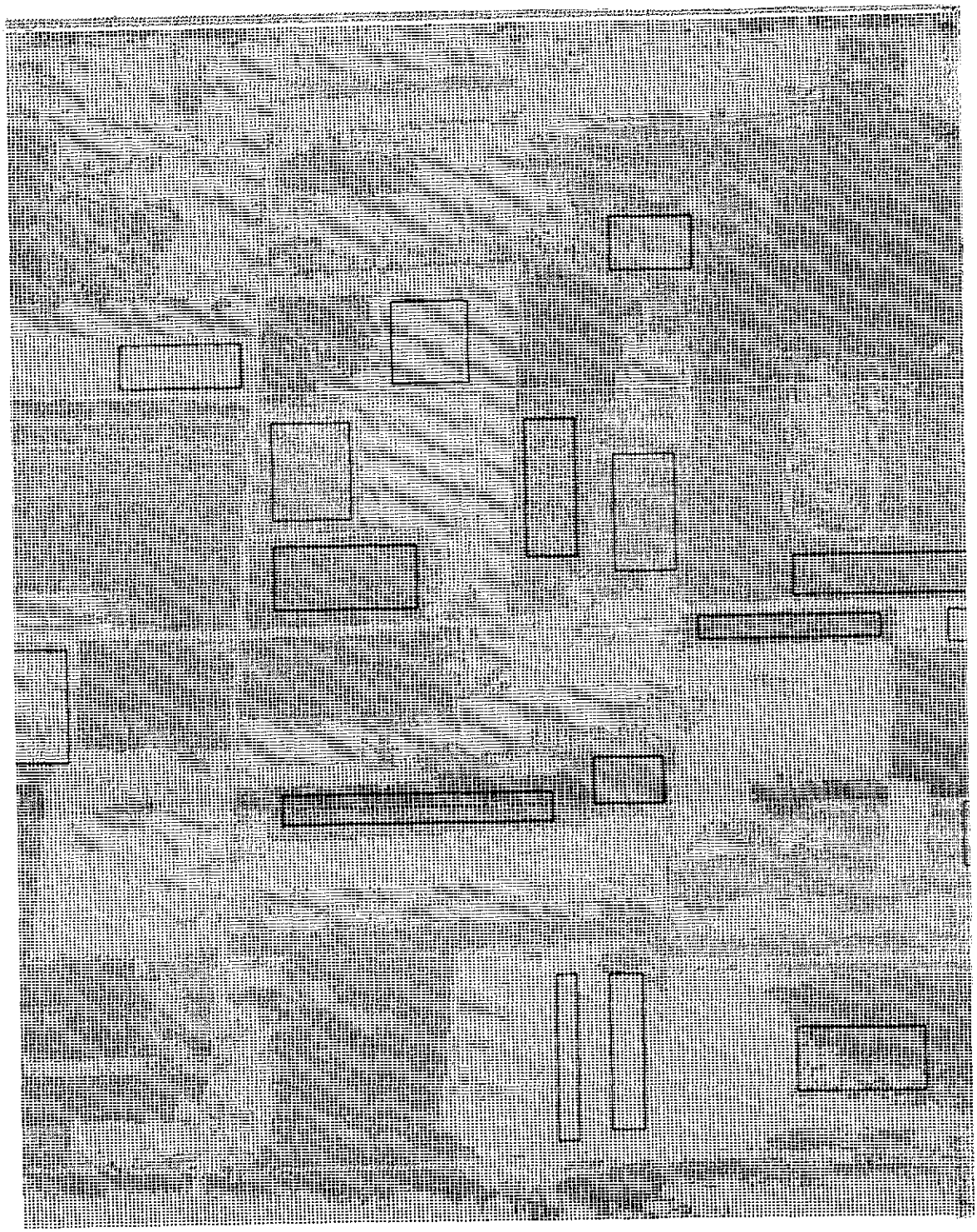


FIGURE 5. KALMAN FILTER RECOGNITION MAP

the data when we use the preprocessing feature of the Kalman filter. We actually used the same signatures, but would have preferred using signatures of the data after it had been preprocessed using our initial values for the polynomials. The average percentage correct recognizing that we measured for the 25 fields was 74.7% when θ_1 was 10^{-7} and 74.3% when θ_1 was 10^{-8}. The other parameters were: $\theta_2 = 0$, $\alpha_1 = 1.0$, and the confidence scale factor was 0.25. Without either preprocessing or updating, the recognition accuracy measured 68.7%.

It would appear that we should have used another method to derive signatures. We also have some doubts about the superiority of the use of the preprocessing feature, compared to our standard preprocessing procedures to improve classification accuracy when only rapid variations in the data are present. It is more realistic to expect improvement when there are gradual changes, especially when the changes are caused by changes in the atmosphere or sun position.

5
FEATURE EXTRACTION OF MULTISPECTRAL DATA *

The data gathered by a multispectral scanner must be processed, for example by recognition processing, before much of the information in the data can be extracted. In carrying out this information extraction it is desirable to reduce processing time to the extent possible. One method for doing this is to reduce the number of channels of data to be processed.

There are several ways to reduce the number of channels to be processed. For example, one could reduce the number of spectral channels in the scanner itself. This method has the disadvantage that, because of changes in atmospheric conditions, changes in the spectral reflectance of crops during various stages of growth, and other variable factors, one does not know in advance of examining the data which spectral channels should be recorded.

A second method is to record all of the channels available, but to process only a subset of channels. This method has been routinely used at ERIM in recent years. The training data are used to decide which channels can be eliminated. We found that 4 to 6 spectral channels provide almost as much recognition information as do all of the channels (usually 12 of them) [20].

Investigation of a third possibility is reported here, namely, to form a subset of linear combinations of the recorded spectral channels [23, 24, 25, 26, 27]. The subset of linear combinations can then be used in recognition processing just as if they were a subset of the original channels. In this case, the training data are examined to determine which linear combinations of channels to use. The subsets of linear combinations would then be calculated for each data point prior to beginning recognition processing.

In many applications, it may be that a subset of 2 or 3 linear combinations will give the same recognition information as if all pure channels were used. If the time required to calculate the linear combinations is small compared to the time savings achieved in recognition resulting from the reduction in the number of channels, the advantage in employing linear combinations may be substantial. With analog data available, the linear combinations can be formed at the time the data are digitized. For data in digital format, it may be convenient to form the linear combinations when the data are converted into a format suitable for recognition, or during the preprocessing operation.

A simple example can illustrate the formation of linear combinations and the performance that is possible. Consider the problem of recognizing one of four possible classes using one linear combination of two-channel data. The data channels are assumed to be independent, each with variance σ^2. The location of the mean values for the four classes is shown in Fig. 6. The calculation of the average probability of misclassification for this geometry was made by assuming that a linear decision rule would be used, that the data were normal and described

* The content of this section has been presented previously. See Ref.[22].

by the means and variance, (i.e., the covariance terms are zero) and that the a priori probabilities of occurrence of each class are equal. The calculations were limited to pairwise evaluations, whereby the probabilities of misclassification for each pair of classes were not affected by the presence of the remaining class. The formation of a linear combination has the effect of projecting the data into a subspace which in this case is just a line.

The calculation results are shown in Fig. 7, where the average probability of misclassification is shown as a function of the angle between the abscissa and projection line that determines the linear combination. The performance that would be obtained by using pure channels is found for angles of 0^{o} and 90^{o}. The lowest average probability of misclassification occurs for a linear combination described by an angle of approximately 15^{o}. Thus for this particular example, a linear combination would be more desirable for recognition than a subset (i.e., a subset of one) of pure channels. Notice that for some linear combinations the average probability of misclassification exceeds 1.0. This occurs because of the pairwise assumption used in the calculation, whereby some points are misclassified in more than one way, and all of the ways are counted in computing the average.

A procedure for finding the best linear combination in this case might be to start with any combination, and compute the average probability of misclassification. Then, the calculation is repeated for a linear combination described by an angle close to the first angle. This procedure is then repeated, always with the angle close to that angle which in the previous computation had provided the lowest average probability of misclassification, until a minimum was found. For this example, where there are two minima, the lowest would be the one that would be found most of the time.

A slight variation of the geometry previously described is shown in Fig. 8. The only difference is the change in the variance, σ. Figure 9 depicts the calculated average probability of misclassification. The averages are all lower than those shown previously, a result of the reduced variance. In addition, there are now 5 minima, rather than 2, so that a minimum seeking technique is more dependent on the starting linear combination. The number of easily detectable minima can be reduced by artificially increasing the variance. This phenomena may lead to an improved minimum seeking technique.

5.1 CHOOSING LINEAR COMBINATIONS

The problem of finding a good method of choosing linear combinations is primarily one of finding a workable algorithm in three distinct steps: (1) develop a measure of performance, (2) develop a minimum seeking technique, and (3) find suitable starting points for initiating the minimum seeking technique. In addition, the algorithm should not require an excessive amount of computational time.

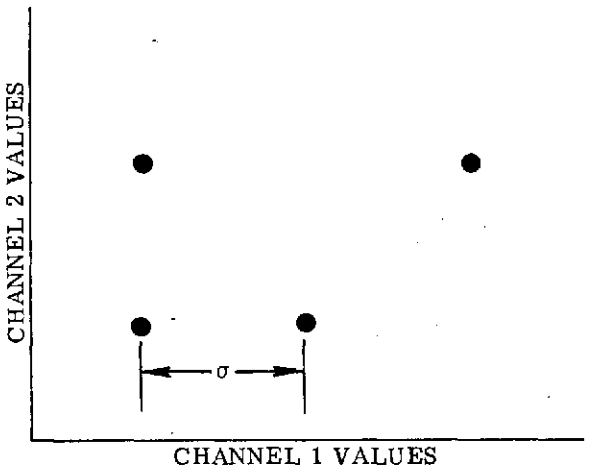


FIGURE 6. LOCATION OF MEANS FOR EXAMPLE WITH LARGE VARIANCE

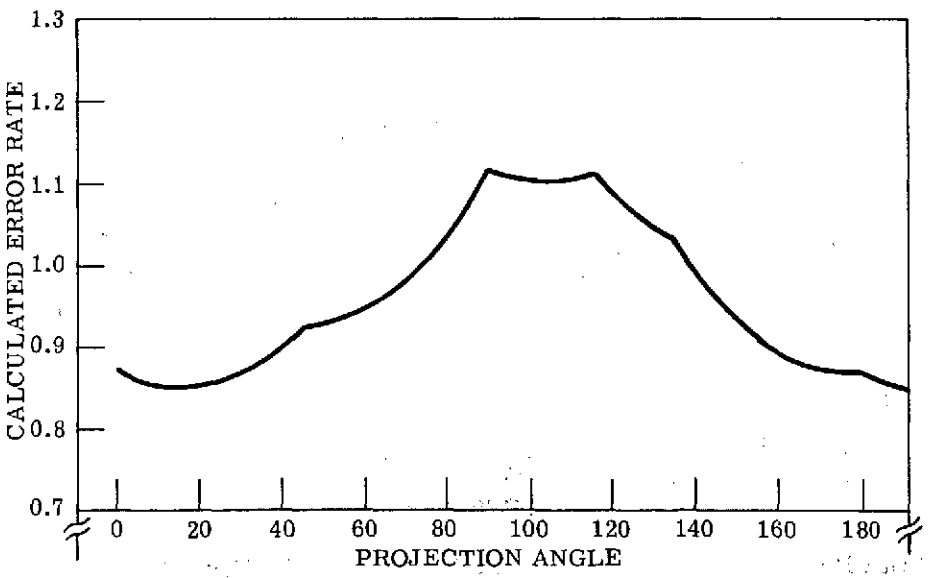


FIGURE 7. CALCULATED RECOGNITION ACCURACY FOR DIFFERENT PROJECTION ANGLES FOR EXAMPLE WITH LARGE VARIANCE

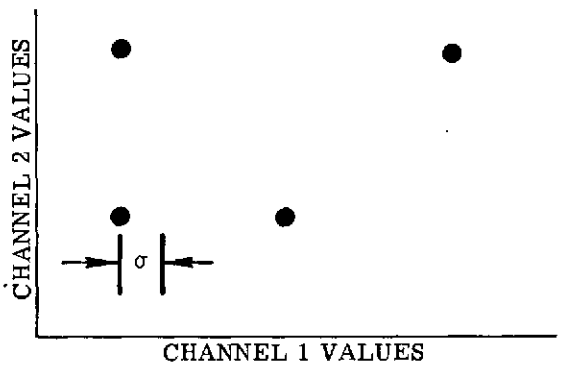


FIGURE 8. LOCATION OF MEANS FOR EXAMPLE WITH SMALL VARIANCE

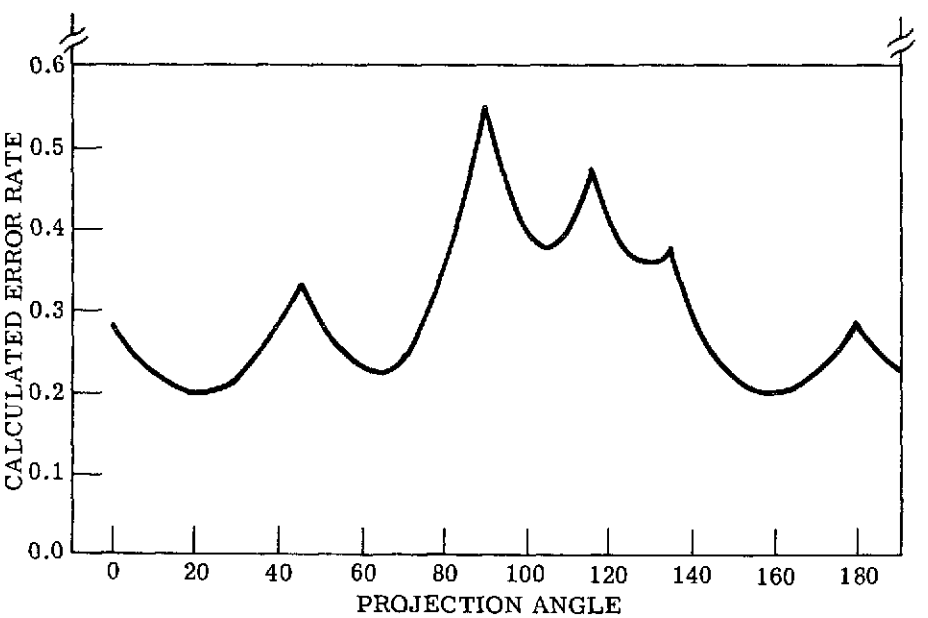


FIGURE 9. CALCULATED RECOGNITION ACCURACY FOR DIFFERENT PROJECTION ANGLES FOR EXAMPLE WITH SMALL VARIANCE

5.1.1 PERFORMANCE MEASURE

The performance measure used to evaluate both subsets of linear combinations and subsets of pure channels can be expressed as:

$$M = \sum_{i,j} \phi\left\{1/2 \left[(\mu_i - \mu_j)^t A^t A\left(\frac{R_i + R_j}{2}\right) A^{t-1} A(\mu_i - \mu_j)\right]^{1/2}\right\} \qquad (65)$$

where the summation is for all signatures, the i-th class is distributed normally with mean vector μ_i and covariance matrix R_i, and $\phi(X)$ is the normal distribution function. The row vectors of the m×n matrix A represent the linear combinations in question (n is the number of pure channels and m is the number of linear combinations). When this expression is used to evaluate a subset of pure channels, the matrix A becomes an array of 0's and 1's. In the past, this expression has been used for this purpose. Notice that if the distance between means, $\mu_i - \mu_j$, is increased the argument of the normal distribution function associated with that pair is increased, resulting in an increase of the expression M. M is thus an approximate measure of the total number of correct classifications. The average probability of misclassification is then approximately 1 minus a term proportional to M.

If R_i and R_j increase, then the average probability of misclassification would be expected to increase. Following through the expression for M we see that this is so. A maximum likelihood linear decision rule (as opposed to a quadratic decision rule) very similar in form to Eq. (65) is now used routinely in this laboratory for recognition processing. The measure M is an accurate measure of the results actually obtained in using this rule with two specific approximations. First, M gives added weight to regions which can be misclassified in more than one way, while a decision rule used on actual data partitions all the data uniquely into separate classes. Second, the linear decision rule actually used on data takes account of both of the covariance matrices $R_i + R_j$, whereas Eq. (65) merely averages them together. Equation (65) is, however, a more accurate measure than others which have been suggested. For example, $(R_i + R_j)/2$ could be replaced by the average of all the covariance matrices in the signature set, rather than using the appropriate pairwise average. Also, the function ϕ could be replaced by its argument, resulting in what has been designated a "divergence criterion." This is clearly an inferior criterion since it continues to force more separation between means even after they are several sigma apart. An advantage of Eq. (65) is that it can be developed directly from the maximum likelihood decision rule, so the approximations used can be enumerated and evaluated. In fact, Eq. (65) is approximately proportional to a constant minus the average probability of misclassification that would be measured.

5.1.2 MINIMUM SEEKING TECHNIQUE

A method has been developed to find a local minimum of a function of several variables by starting at a point and following a path of steepest descent by steps of variable but controllable size. Both the local gradient and the local curvature are used to estimate the path of steepest descent.

There is one additional problem concerning the determination of A that should be mentioned. If A is an m×n matrix, there are mn components to be determined. This number of components can be reduced to m(n - m) by the choice of a suitable canonical form of A. A canonical form is possible because the value of M obtained for any A is not changed if PA is substituted for A, where P is any nonsingular matrix. We actually use PA, where P is chosen to scale the average covariance matrix and the mean vectors of the materials to our data format. The canonical form we chose is:

$$A = \begin{bmatrix} & \tan\theta_{11} \tan\theta_{12} \cdots \tan\theta_{1(n-m)} \\ & \tan\theta_{21} \\ I_m & \\ & \cdot \\ & \cdot \\ & \cdot \\ & \tan\theta_{m,1} \end{bmatrix} \qquad (66)$$

where I_m is the identity matrix with rank m. (For a specific example see Table 8, where m = 3, n = 10, and the 10 pure channels have been rearranged in order of the wavelength.)

The canonical form with the θ_{ij} has two advantages. The first is that, in general, a minimum number of unknown scalars must be found. The second advantage is that the minimization process can be accomplished by varying the θ_{ij} with a nearly uniform step size. It is not necessary to have large jumps in the values of the unknown scalars, which occur if the $\tan\theta_{ij}$ are considered to be the unknown scalars.

5.1.3 FIND STARTING POINTS

Finding starting points, the third step, is more difficult. The following are suggested starting points.

Best Subset of Channels Starting Point

Each individual channel can be thought of as a linear combination of channels. (The vector representing this combination has a 1 in the appropriate coordinate and 0's elsewhere.) Therefore, a subset of m channels can be thought of as a set of m linear combinations. Since the

number of subsets of m channels can be large, rather than check through all of them to find the best one, we use a stepwise procedure to find a "good" one. This stepwise procedure successively adds the one channel which gives the lowest average probability of misclassification when used with the channels already selected. The linear combinations represented by this subset are then used as a starting point.

Norm-Squared Starting Point

By replacing each covariance matrix by the average of all of them, the problem is reduced to minimizing the function.

$$M(P) = \sum_i \phi(1/2 \, || Pw_i ||) \tag{67}$$

where the w_i are a fixed set of vectors and P ranges over all orthogonal projections of rank m. The number of vectors w_i is the total number of pairwise combinations of signatures. Each projection P corresponds in a simple fashion to a matrix A in the original formulation (see [23]). The projection P which maximizes

$$\sum_i || Pw_i ||^2 \tag{68}$$

is found analytically and the corresponding A is used as a starting point.

Principal Eigenvector Starting Point

First calculate the average of all the covariance matrices. Then transform the data so that the average covariance matrix is the identity matrix. Let $N(\mu, R)$ be the distribution of all the transformed data in the training area lumped together. This distribution can be calculated from the distributions of the various materials if we can estimate the frequency of occurrence of each material. The starting point A is then taken as the matrix whose row vectors are the m orthogonal eigenvectors corresponding to the m largest eigenvalues of the covariance matrix R.

Clustered Starting Point

This method is based on the fact if there are only two signatures and we are using linear discrimination, then there always exists a single linear combination channel which distinguishes exactly as well as all n channels no matter what the value of n. If there are many signatures, then for each pair S_i, $S_j (i \neq j)$, v_{ij} can be the unit vector corresponding to this best single linear combination. In general, the number of vectors $v_{i,j}$, will be greater than m. The $v_{i,j}$ are then clustered into m clusters. For each cluster C_k, $k = 1, \ldots, m$, a weighted average, w_k, of the $v_{i,j}$ in that cluster is computed. The starting point A is formed from the w_k as row vectors. The weights can be made to reflect the sensitivity of the recognition accuracy to the decision rule.

5.2 EXPERIMENTAL RESULTS

We compared subsets of linear combinations with subsets of pure channels. The data used were from one of the sets previously employed to test our linear decision rule [1]. We chose this particular set because of the difficulty we have noticed in obtaining satisfactory recognition with it. We felt that with relatively poor recognition accuracy, the test results would represent greater statistical accuracy. If only a few data points were incorrectly recognized, the test results would be too dependent on those few points.

The test procedure we used was to first select data that corresponded to 20 training fields. From these fields we developed statistics (mean and covariance) for each of the 7 classes of materials. The statistics or signatures were then used to develop the decision rules which were applied to data that corresponded to 23 test fields different from the training fields. We then found the average correct recognition for each field, and then the average for each material. Finally we averaged recognition accuracies for the materials to obtain an average recognition accuracy for the data set. The computer programs were merely functional, not optimized for minimum computation time, so meaningful comparisons of computation times were not made.

The material classes consisted of bare soil and six vegetative species: alfalfa, barley, lettuce, sugar, safflower, and rye. The bare soil data tended to be atypical, because three or more pure channels of data provided almost perfect recognition, whereas all of the subsets of 3 linear combinations of channels provided reduced accuracy. Note that the various subsets of linear combinations were chosen to optimize over all the species; therefore, it is not surprising that they did less well for one of them. There is some evidence that one infrared channel or ratio of channels can be used to separate vegetative and non-vegetative materials. Thus,for some applications of layered or sequential classifiers, bare soil may not be considered as a class to be recognized when discriminating among vegetation types. However, for this study, we retained bare soil as a class.

The test results are shown in Fig. 10. The entire bar indicates the average recognition accuracy obtained for the 7 classes. The unshaded portion indicates accuracies for the 6 vegetative classes. Note that the recognition accuracy for the subset of 3 linear combinations was better than that obtained when subsets of either 3 or 4 best pure channels were used. In fact, the accuracy approached that obtained for all 10 channels, especially when only the 6 vegetative materials are considered.

Figure 10 shows the average recognition accuracy obtained for one subset of 3 linear combinations only. We actually tested three subsets of 3 linear combinations. Two of the subsets resulted from minimizing our measure function with two different starting points (the first of these was used for Fig. 10) and the third subset was an unweighted addition of channels. We obtained approximately the same average recognition accuracy for each of the subsets of linear combinations.

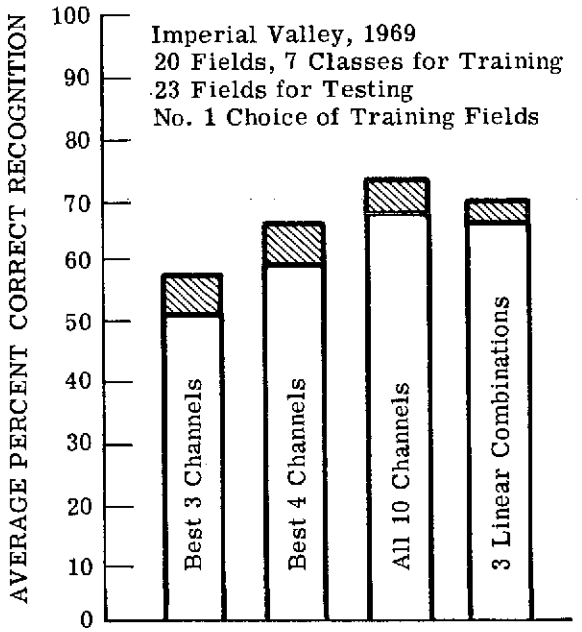


FIGURE 10. COMPARISON OF LINEAR COMBINATIONS WITH SUBSETS OF CHANNELS FOR TEST FIELDS

In Table 9, the average recognition accuracies we obtained from the 3 subsets of 3 linear combinations are compared with each other and with subsets of pure channels. Note the correspondence with the predicted accuracies, especially for subsets of pure channels without the bare soil class. The accuracies were predicted from Eq. (65) with all signatures, including bare soil, being used.

The first subset of linear combinations is shown in Table 8, each row represents one combination. This matrix is not determined uniquely, because premultiplication by any nonsingular matrix results in a new set of linear combinations which would reduce identical recognition performance. The starting point for this set was the subset of 3 pure channels that we used for comparison.

TABLE 8. MATRIX DESCRIBING 3 LINEAR CHANNELS WHEN THE STARTING POINT IS THE BEST SUBSET OF 3 CHANNELS.
Each row contains a set of weights that determines one of the linear combinations.

1	0.29	0.45	-0.74	-0.42	-0.06	0.07	0	0	-0.20
0	0.23	0.65	-0.02	0.60	0.76	0.82	1	0	-0.01
0	0.18	0.57	-0.34	0.20	0.31	0.11	0	1	-0.09

TABLE 9. PERCENTAGE RECOGNITION ACCURACY OBTAINED BY USING LINEAR COMBINATIONS, SUBSETS OF CHANNELS, AND ANALYTIC PREDICTION

	LINEAR COMBINATIONS				SUBSETS		
	No. 1	No. 2	No. 3		3	4	10
	Recognition (%)				Recognition (%)		
Without soil	66	66	64		51	59	68
With soil	67	70	69		58	65	73
Predicted	71	75			51	62	70

6
CONCLUSIONS AND RECOMMENDATIONS

A general linear theory can be used for adapting the means of class signatures. The theory can accommodate the use of known additive and multiplicative changes to the means, use of uncertain decision results or proportion estimates, a method to use auxiliary ground truth, and a method of adapting scan angle corrections. The general theory, which is essentially a decision-directed Kalman filter, includes as special cases the adapting algorithms which had been derived empirically and tested during last year's contract.

Two limiting features of the Kalman filter can be avoided simultaneously. These are: (1) the requirement that first- and second-order statistics of the variations of the means be known, and (2) large memory and computation requirements are generally associated with a Kalman filter containing many states. Because the Kalman filter model is only an approximation to a description of remotely sensed data, the use of the more accurate statistics may not produce a significant improvement in the accuracy of the mean estimates.

The development and testing program was devoted to finding simple, practical methods, rather than to the full utilization of the generality of the theory. The test results provided confirmation of the usefulness of the Kalman filter model. It was shown that processing accuracy could be improved over that obtained previously with empirically derived algorithms, as well as that obtained when the means were held constant.

We conclude from these limited tests that the Kalman filter algorithm can improve classification accuracy in two ways: The updated means can be made to follow variations in the data caused by inter- and intra-field changes in the ground covers. Or, the updated means can be restricted to show variations which would follow slow changes resulting from atmospheric variations, varying sun position, or varying ground cover reflectances in the data.

The ability of the Kalman filter to adapt to rapid changes in the data may have an important operational consequence. It should be possible to better delineate fields and field boundaries. This may improve the accuracy of estimating acreage, e.g., specific agricultural crops, as well as increasing the probability of a correct decision of the ground cover on a given field.

As a result of the test results, one conclusion must be that adaptive processing is a processing method capable of improving recognition accuracies, at least for some data sets. Additional testing should indicate the general usefulness of the technique and the parameters to be used. There are indications that the linear, rather than quadratic, decision rule would increase classification accuracy and decrease processing time. We believe that the filter should be implemented so that auxiliary training fields can be used to decrease the probability of capture.

From observing the test results, we have concluded that the Kalman filter model should be changed so that the means adapt to slow the rapid changes in the data simultaneously. If this change is made, the updating rate should increase without increasing the probability of capture.

We did not try to update the signature means while estimating proportions of unresolved objects. In addition, we did not try using any auxiliary information. The sun sensor in the ERIM M-7 multispectral scanner is one source of auxiliary information presently available.

We believe that adaptive processing is a useful method of classifying multispectral data. Additional testing with other data sets should be performed so that the capabilities and limitations can be better understood and the method can come to be used routinely. The Kalman filter processor can fulfill many functions simultaneously. In fact, it appears that many practical adaptive algorithms can be shown to be equivalent to a Kalman filter.

Finally, we conclude from our study of feature extraction that the use of linear combinations may be a feasible method of spectral feature extraction to reduce overall processing time. The tests should be extended to include more data sets and different starting points.

Appendix A
BASIC KALMAN FILTER

The Kalman filter has been known, used, and described for some years. The development that follows [28] is intended as a review and to show one method of development that can be used to obtain all of the Kalman filter variations that are described in this report.

The basic assumption for the theory is that a process can be described by the two equations:

$$x_k = x_{k-1} + w_{k-1} \tag{A.1}$$

$$z_k = H_k x_k + v_k \tag{A.2}$$

where x_k is an nm×1 state vector defined at time t_k; z_k is an n×1 measurement vector, H_k is an n×nm transition matrix; and w_k and v_k are random vectors with dimensions nm × 1 and n × 1 respectively. We further assume that we have a Markov process, i.e.,

$$E(w_k) = 0, \quad k = 1, \ldots \tag{A.3}$$

$$E\left(w_k w_j^t\right) = \delta_{kj} Q_k, \quad j, k = 1, 2, \ldots \tag{A.4}$$

$$E(v_k) = 0, \quad k = 1, \ldots \tag{A.5}$$

$$E\left(v_k v_j^t\right) = \delta_{kj} R_k, \quad j, k = 1, 2, \ldots \tag{A.6}$$

$$E\left(w_k v_j^t\right) = 0, \quad j, k = 1, 2, \ldots \tag{A.7}$$

We wish to estimate x_k with $\hat{x}_k$, knowing $z_1, \ldots, z_k$, and we wish our estimate to be optimum in the sense that we minimize trace P_k, where

$$P_k = E\left\{(\hat{x}_k - x_k)(\hat{x}_k - x_k)^t\right\} \tag{A.8}$$

We shall estimate x_k with the equation

$$\hat{x}_k = \hat{x}_{k-1} + K_k(z_k - H_k \hat{x}_{k-1}) \tag{A.9}$$

where K_k is to be determined. Equation(A.9)is not the most general form of linear estimators, but it can be shown that we shall obtain the optimum estimate. Of course we have restricted our estimate to be a linear function of the measurements, but our estimate is actually optimum even without this restriction if v_k and w_k are Gaussian random vectors.

We wish to find the K_k that minimizes the trace of

$$P_k = E\left(\tilde{x}_k \tilde{x}_k^t\right) \tag{A.10}$$

where

$$\tilde{x}_k = \hat{x}_k - x_k \tag{A.11}$$

$$= \hat{x}_{k-1} + K_k(H_k x_k + v_k - H_k \hat{x}_{k-1}) - x_k \tag{A.12}$$

which we obtain by combining Eqs. (A.2), (A.9), and (A.11). We now use Eq. (A.1) and get

$$\tilde{x}_k = \hat{x}_{k-1} + K_k(H_k x_{k-1} + H_k w_{k-1} + v_k - H_k \hat{x}_{k-1}) - x_{k-1} - w_{k-1}$$

$$= (I - K_k H_k)(\tilde{x}_{k-1} - w_{k-1}) + K_k v_k \tag{A.13}$$

We next form P_k, using Eqs. (A.7), (A.10), and (A.13).

$$P_k = E\left[(I - K_k H_k)(\tilde{x}_{k-1} - w_{k-1})(\tilde{x}_{k-1} - w_{k-1})^t (I - K_k H_k)^t\right] + E\left(K_k v_k v_k^t K_k^t\right) \tag{A.14}$$

$$= (I - K_k H_k) P_k' (I - K_k H_k)^t + K_k R_k K_k^t \tag{A.15}$$

where

$$P_k' = P_{k-1} + Q_{k-1} \tag{A.16}$$

We now complete the square in K_k, which results in

$$P_k = (K_k - A)B(K_k - A)^t - ABA^t + P_k' \tag{A.17}$$

where

$$A = P_k' H_k^t \left(H_k P_k' H_k^t + R_k\right)^{-1} \tag{A.18}$$

$$B = H_k P_k' H_k^t + R_k \tag{A.19}$$

We can now minimize trace P_k by letting

$$K_k = A = P_k' H_k^t \left(H_k P_k' H_k^t + R_k\right)^{-1} \tag{A.20}$$

With this value of K_k,

$$P_k = P_k' - K_k H_k P_k' \tag{A.21}$$

We have now completed the development of the basic Kalman filter. The filter is described by Eqs. (A.9), (A.16), (A.20), and (A.21).

Appendix B
A KALMAN FILTER FOR AUXILIARY GROUND TRUTH

The Kalman filter will be derived following the procedures used in Appendix A. We first define the problem:

$$x_k = x_{k-1} + w_{k-1} \tag{B.1}$$

$$Z_k = H_k X_k + v_k \tag{B.2}$$

$$\hat{x}_k = \hat{x}_{k-1} + K_k(Z_k - H_k\hat{x}_{k-1}) + \sum_{i=1}^{p} J_{ki}H_{Ni}(x_{Ni} - \hat{x}_{k-1}) \tag{B.3}$$

$$\tilde{x}_k = \hat{x}_k - x_k \tag{B.4}$$

$$P_k = E\left(\tilde{x}_k \tilde{x}_k^t\right) \tag{B.5}$$

$$P'_k = P_{k-1} + Q_{k-1} \tag{B.6}$$

We wish to find K_k and J_{ki} that minimize trace P_k. The number of ground truth fields is p. We assume that at point N_i we know $H_{Ni}x_{Ni}$ (the means computed from the auxiliary ground truth). The first step is to combine Eqs. (B.3) and (B.4); using (B.1) and (B.2):

$$\begin{aligned}\tilde{x}_k &= \hat{x}_k - x_k - \hat{x}_{k-1} + K_k(H_k x_{k-1} + H_k w_{k-1} + v_k - H_k\hat{x}_{k-1} \\ &\quad + \sum_{i=1}^{p} J_{ki}H_{Ni}(x_{Ni} - \hat{x}_{k-1}) - K_{k-1} - w_{k-1} \\ &= (I - K_k H_k)(\tilde{x}_{k-1} - w_{k-1}) + K_k v_k \\ &\quad + \sum_{i=1}^{p} J_{ki}H_{Ni}(x_{Ni} - \hat{x}_{k-1})\end{aligned} \tag{B.7}$$

Before we compute Eq. (B.5), we shall compute two expectations. We first note that

$$\begin{aligned}x_{Ni} - \hat{x}_{k-1} &= x_{k-1} + w_{k-1} + w_k + \cdots + w_{Ni-1} - \hat{x}_{k-1} \\ &= w_{k-1} + \cdots + w_{Ni-1} - \tilde{x}_{k-1}\end{aligned} \tag{B.8}$$

If we use Eq. (B.8), we find that, for $i \geq j$

$$E\left[(x_{Ni} - \hat{x}_{k-1})(x_{Nj} - \hat{x}_{k-1})^t\right] = P_{k-1} + Q_{k-1} + \cdots + Q_{Nj-1} = P'_k + (N_j - k)Q \tag{B.9}$$

when $Q_i = Q$ for all i. Also,

$$E\left[(\tilde{x}_{k-1} - w_{k-1})(x_{Ni} - \hat{x}_{k-1})^t\right] = E[(\tilde{x}_{k-1} - w_{k-1})(w_{k-1} + \ldots + w_{Ni-1} - \tilde{x}_{k-1})] = -P'_k \qquad (B.10)$$

We are now at the point where we can compute Eq. (B.5).

$$P_k = E\left\{\left[(I - K_kH_k)(\tilde{x}_{k-1} - w_{k-1}) + K_kv_k + \sum_{i=1}^{p} J_{ki}H_{Ni}(x_{Ni} - \hat{x}_{k-1})\right]\left[(I - K_kH_k)(\tilde{x}_{k-1} - w_{k-1})\right.\right.$$

$$\left.\left. + K_kv_k + \sum_{i=1}^{p} J_{kj}H_{Nj}(x_{Nj} - \hat{x}_{k-1})\right]^t\right\} = (I - K_kH_k)P'_k(I - K_kH_k)^t + K_kR_kK_k^t$$

$$- \sum_{j=1}^{P}(I - K_kH_k)P'_kH_{Nj}^tJ_{kj}^t - \sum_{i=1}^{p} J_{ki}H_{Ni}P'_k(I - K_kH_k)^t$$

$$+ \sum_{i,j=1}^{P} J_{ki}H_{Ni}[P'_k + (N_{i,j-k})Q]H_{Nj}^tJ_{kj}^t \qquad (B.11)$$

where

$$N_{i,j} = N_i,\ i \le j$$
$$= N_j,\ j \le i \qquad (B.12)$$

We shall simplify Eq. (B. 11) before finding K_k and J_{ki}. The last term of Eq. (B.11) is

$$T_1 = \sum_{i,j=1}^{p} J_{ki}H_{Ni}[P'_k + (N_{i,j-k})Q]H_{Nj}^tJ_{kj}^t$$

$$= \sum_{i\le j}^{p} J_{ki}H_{Ni}[P'_k + (N_j - k)Q]H_{Nj}^tJ_{kj}^t \qquad (B.13)$$

$$+ \sum_{i\ge j}^{p} J_{ki}H_{Ni}[P'_k + (N_j - k)Q]H_{Nj}^tJ_{kj}^t$$

The second part of Eq. (B.13) can be written

$$\sum_{i>j}^{p} J_{ki}\left\{H_{Nj}[P'_k + (N_j - k)Q]H_{Ni}^t\right\}^tJ_{kj}^t$$

If we define

$$F_{ijk} = H_{Ni}[P'_k + (N_i - k)Q]H_{Nj}^t$$

Eq. (B.13) becomes

$$T_1 = \sum_{i=j}^{p} J_{ki} F_{ijk} J_{kj}^t + \sum_{i>j}^{p} J_{ki} F_{jik}^t J_{ki}^t \tag{B.14}$$

Eq. (B.14) has the form of a vector-matrix-vector product, although each term is a matrix. When we use partitioned matrices, we have

$$T_1 = J_k F J_k^t \tag{B.15}$$

where

$$J_k = (J_{k1} \cdots J_{kp}) \tag{B.16}$$

$$F = \begin{pmatrix} F_{11k} & F_{12k} \\ F_{12k}^t & F_{22k} \\ \cdot & \\ \cdot & \\ \cdot & \end{pmatrix} \tag{B.17}$$

Two other terms of Eq. (B.11) should be simplified:

$$T_2 = \sum_{j=1}^{p} (I - K_k H_k) P_k' H_{Nj}^t J_{kj}^t \tag{B.18}$$

$$T_3 = \sum_{i=1}^{p} J_{ki} H_{Ni} P_k' (I - K_k H_k)^t \tag{B.19}$$

We first note that $T_3 = T_2^t$. We also note that the terms in Eq. (B.18) that depend on the index j have the form of an inner product. Thus we write

$$T_2 = (I - K_k H_k) P_k' H_N^t H_N^t J_k^t \tag{B.20}$$

where J_k is defined in Eq. (B.16) and

$$H_N = \begin{pmatrix} H_{N1} \\ \cdot \\ \cdot \\ \cdot \\ H_{Np} \end{pmatrix} \tag{B.21}$$

This completes the simplification of Eq. (B.11), which can now be put into the form

$$P_k = (I - K_k H_k)P'_k(I - K_k H_k)^t + K_k R_k K_k^t - (I - K_k H_k)P'_k H_N^t J_k^t$$

$$- J_k H_N P'_k (I - K_k H_k)^t + J_k F J_k^t \tag{B.22}$$

The next step is to find J_k that minimizes P_k. We shall use the identity

$$J_k F J_k^t - J_k F A^t - A F J_k^t = (J_k - A)F(J_k - A)^t - A F A^t \tag{B.23}$$

with

$$A = (I - K_k H_k)P'_k H_N^t F_k^{-1} \tag{B.24}$$

Equation (B.22) becomes

$$P_k = (I - K_k H_k)P'_k(I - K_k H_k)^t + K_k R_k K_k^t + (J_k - A)F(J_k - A)^t - A F A^t \tag{B.25}$$

Thus, independently of K_k we choose

$$J_k = A = (I - K_k H_k)P'_k H_N^t F^{-1} \tag{B.26}$$

and have

$$P_k = (I - K_k H_k)P'_k(I - K_k H_k)^t + K_k R_k K_k^t - (I - K_k H_k)P'_k H_N^t F^{-1} H_N P'_k (I - K_k H_k)^t$$

$$= (I - K_k H_k)P''_k(I - K_k H_k)^t + K_k R_k K_k^t \tag{B.27}$$

where

$$P''_k = P'_k - P'_k H_N^t F^{-1} H_N P'_k \tag{B.28}$$

The minimization of Eq. (B.27) with a choice of K_k is in Appendix A.

$$K_k = P''_k H_k^t (H_k P''_k H_k^t + P_k)^{-1} \tag{B.29}$$

$$P_k = P''_k - K_k H_k P''_k \tag{B.30}$$

This completes the development.

REFERENCES

1. Crane, R. B.,W. M. Malila, and W. Richardson, Suitability of the Normal Density Assumption for Processing Multispectral Scanner Data, IEEE Transactions on Geoscience Electronics, Vol. GE 10 #4, Oct. 1972.

2. Horwitz, H., P. Hyde, and W. Richardson, Improvements in Estimating Proportions of Objects from Multispectral Data, Report No. 190100-25-T, Environmental Research Institute of Michigan, Ann Arbor, (In Press).

3. Crane, R. B. Preprocessing Techniques to Reduce Atmospheric and Sensor Variability in Multispectral Scanner Data, Seventh International Symposium on Remote Sensing of Environment, Report No. 10259-1-X, Willow Run Laboratories of the Institute of Science and Technology, The University of Michigan, Ann Arbor, May 1971.

4. Kriegler, F. J., R. F. Nalepka, W. M. Malila and W. Richardson, Preprocessing Transformations and Their Effects on Multispectral Recognition, Sixth International Symposium on Remote Sensing of Environment, Report No. 31069-2-X, Willow Run Laboratories of the Institute of Science and Technology, The University of Michigan, Ann Arbor, October 1969.

5. Nalepka, R. F. and J. P. Morgenstern, Signature Extension Techniques Applied to Multispectral Scanner Data, Eighth International Symposium on Remote Sensing of Environment, Report No. 195600-1-X, Willow Run Laboratories of the Institute of Science and Technology, The University of Michigan, Ann Arbor, October 1972.

6. Malila, W., R. Hieber, and J. Sarno, Analysis of Multispectral Signatures and Investigation of Multi-Aspect Remote Sensing Techniques, Report No. 190100-27-T, Environmental Research Institute of Michigan, Ann Arbor, (In Press).

7. Vincent, R., G. Thomas, and R. Nalepka, Signature Extension Study, Report No. 190100-26-T, Environmental Research Institute of Michigan, Ann Arbor, (In Press).

8. Turner, R. E., W. M. Malila and R. F. Nalepka, Importance of Atmospheric Scattering in Remote Sensing or Everything You've Always Wanted to Know About Atmospheric Scattering But Were Afraid to Ask, Seventh International Symposium on Remote Sensing of Environment, Report No. 10259-1-X, Willow Run Laboratories of the Institute of Science and Technology, The University of Michigan, Ann Arbor, May 1971.

9. Turner, R. E. and Spencer, M. M., Atmospheric Model for Correction of Spacecraft Data, Eighth International Symposium on Remote Sensing of Environment, Report No. 195600-1-X, Willow Run Laboratories of the Institute of Science and Technology, The University of Michigan, Ann Arbor, October 1972.

10. Vincent, R., New Theoretical Methods in Ratio-Imaging Techniques Associated with the NASA Earth Resources Spectral Information System, Report No. 190100-30-T, Environmental Research Institute of Michigan, Ann Arbor, (In Press).

11. Marshall, R. E., F. J. Kriegler and W. Richardson, Adaptive Multispectral Recognition of Wheat, Tenth Symposium on Adaptive Processes, Miami Beach, December 1971.

12. Kriegler, F. J., R. E. Marshall, H. H. Horwitz and M. F. Gordon, Adaptive Multispectral Recognition of Agricultural Crops, Eighth International Symposium on Remote Sensing of Environment, Willow Run Laboratories of the Institute of Science and Technology, The University of Michigan, Ann Arbor, October 1972.

13. Bucy, R. S. and P. D. Joseph, Filtering for Stochastic Processes with Applications to Guidance, Interscience Publishers, 1968.

14. Sorenson, H. W. Kalman Filtering Techniques, Advances in Control Systems, Vol. 3, (C. T. Leondes, ed.), Acad. Press, 1966.

15. Richardson, W., A Study of Some Nine-Element Decision Rules, Report No. 190100-32-T, Environmental Research Insitute of Michigan, Ann Arbor, (In Press).

16. Horwitz, H. H., R. F. Nalepka, and P. D. Hyde, Estimating the Proportions of Objects Within a Single Resolution Element of a Multispectral Scanner, Seventh International Symposium on Remote Sensing of Environment, Report No. 10259--1-X, Willow Run Laboratories of the Institute of Science and Technology, The University of Michigan, Ann Arbor, May 1971.

17. Nalepka, R. F. and P. D. Hyde, Classifying Unresolved Objects from Simulated Space Data, Eighth International Symposium on Remote Sensing of Environment, Report No. 195600-1-X, Willow Run Laboratories of the Institute of Science and Technology, The University of Michigan, Ann Arbor, October 1972.

18. Crane, R. B. and P. D. Hyde, Signature Estimation from Satellite Multispectral Scanner Data, Remote Sensing of Earth Resources, Vol. 2, (F. Shahrokhis, ed.), The University of Tennessee Space Institute, Tullahoma, Tennessee, 1973.

19. Turner, R. E. Radiative Transfer in Real Atmospheres, Report No. 190100-24-T, Environmental Research Institute of Michigan, Ann Arbor, (In Press).

20. Crane, R. B. and W. Richardson, Performance Evaluation of Multispectral Scanner Classification Methods, Eighth International Symposium on Remote Sensing of Environment, Willow Run Laboratories of the Institute of Science and Technology, The University of Michigan, Ann Arbor, October 1972.

21. Crane, R. B. and W. Richardson, Rapid Processing of Multispectral Scanner Data Using Linear Techniques, Remote Sensing of Earth Resources, Volume 1, (F. Shahrokhis, ed.), The University of Tennessee Space Institute, Tullahoma, Tennessee, 1972.

22. Crane, R. B., T. Cummins and J. F. Reyer, Feature Extraction of Multispectral Data, Machine Processing of Remotely Sensed Data Conference Proceedings, Oct. 16-18, 1973, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.

23. Crane, R. B., W. Richardson, R. Hieber and W. A. Malila, A Study of Techniques for Processing Multispectral Scanner Data, Report No. 31650-155-T, Environmental Research Institute of Michigan, Ann Arbor, September 1973.

24. Crane, R. B., Linear Combinations, Quarterly Supporting Research and Technology Review, Johnson Space Center, Houston, 26-28 September 1972, Report No. 31650-147-L, Environmental Research Institute of Michigan, Ann Arbor, October 1972.

25. Hsia, W. S. and J. P. de Figueiredo, Optimal Feature Extraction - The Two-Class Case, IOSA-275-025-010, Institute for Computer Services and Applications, Rice University, Houston, Texas, May 1973.

26. Jegewski, D. J., Optimal Feature Extraction by a Linear Transformation, Revision 1, MSC Internal Note No. 73-FM-19, 5 March 1973.

27. Quirein, J. A., An Interactive Approach to the Feature Selection Classification Problem, Technical Note 99900-H019-R0-00, TRW Systems, Earth Resources Technology, Houston Operations, December 1972.

28. Root, W. L., Computer, Information and Control Engineering, College of Engineering, The University of Michigan, Ann Arbor, Private Communication to R. B. Crane, Environmental Research Institute of Michigan, Ann Arbor, 1973.

DISTRIBUTION LIST

NASA/Johnson Space Center
Earth Observations Division
Houston, Texas 77058

ATTN: Dr. A Potter/TF3 (4)
ATTN: Mr. Robert MacDonald (1)
ATTN: Mr. B. Baker/TF3 (8)
ATTN: Mr. R. Dean Bretton/TF53
Chief Earth Resources Research Data Facility (8)
ATTN: Mr. A. H. Watkins/HA (1)

NASA/Johnson Space Center
Facility & Laboratory Support Branch
Houston, Texas 77058

ATTN: Mr. D. Riley BB631/B4 (1)

NASA/Johnson Space Center
Computation & Flight Support
Houston, Texas 77058

ATTN: Mr. Eugene Davis/FA (1)

NASA Headquarters
Washington, D.C. 20546

ATTN: Mr. C. W. Mathews (1)

U.S. Department of Agriculture
Agricultural Research Service
Washington, D.C. 20242

ATTN: Dr. Robert Miller (1)

U.S. Department of Agriculture
Soil & Water Conservation Research Division
P.O. Box 267
Weslaco, Texas 78596

ATTN: Dr. Craig Wiegand (1)

U.S. Department of Interior
Geological Survey
Washington, D.C. 20244

ATTN: Dr. James R. Anderson (1)

Director, Remote Sensing Institute
South Dakota State University
Agriculture Engineering Building
Brookings, South Dakota 57006

ATTN: Mr. Victor I. Myers (1)

U.S. Department of Interior
Fish & Wildlife Service
Bureau of Sport Fisheries & Wildlife
Northern Prairie Wildlife Research Center
Jamestown, North Dakota 58401

ATTN: Mr. Harvey K. Nelson (1)

U.S. Department of Agriculture
Forest Service
240 W. Prospect Street
Fort Collins, Colorado 80521

ATTN: Dr. Richard Driscoll (1)

U.S. Department of Interior
Geological Survey
Water Resources Division
500 Zack Street
Tampa, Florida 33602

ATTN: Mr. A. E. Coker (1)

U.S. Department of Interior
Director, EROS Program
Washington, D.C. 20244

ATTN: Mr. J. M. Denoyer (1)

Earth Resources Laboratory, GS
Mississippi Test Facility
Bay St. Louis, Mississippi 39520

ATTN: Mr. R. O. Piland, Director (1)
ATTN: Mr. Sid Whitley (1)

U.S. Department of Interior
Geological Survey
GSA Building, Room 5213
Washington, D.C. 20242

ATTN: Mr. W. A. Fischer (1)

NASA Wallops
Wallops Station, Virginia 23337

ATTN: Mr. James Bettle (1)

Purdue University
Purdue Industrial Research Park
1200 Potter
West Lafayette, Indiana 47906

ATTN: Dr. David Landgrebe (1)
ATTN: Dr. Philip Swain (1)
ATTN: Mr. Terry Phillips (1)

U.S. Department of Interior
EROS Office
Washington, D.C. 20242

ATTN: Dr. Raymond W. Fary (1)

U.S. Department of Interior
Geological Survey
801 19th Street, N.W.
Washington, D.C. 20242

ATTN: Mr. Charles Withington (1)

U.S. Department of Interior
Geological Survey
801 19th Street, N.W.
Washington, D.C. 20242

ATTN: Mr. M. Deutsch (1)

U.S. Geological Survey
801 19th Street, N.W., Room 1030
Washington, D.C. 20242

ATTN: Dr. Jules D. Friedman (1)

U.S. Department of Interior
Geological Survey
Federal Center
Denver, Colorado 80225

ATTN: Dr. Harry W. Smedes (1)

U.S. Department of Interior
Geological Survey
Water Resources Division
901 S. Miami Ave.
Miami, Florida 33130

ATTN: Mr. Aaron L. Higer (1)

University of California
School of Forestry
Berkeley, California 94720

ATTN: Dr. Robert Colwell (1)

School of Agriculture
Range Management
Oregon State University
Corvallis, Oregon 97331

ATTN: Dr. Charles E. Poulton (1)

U.S. Department of Interior
EROS Office
Washington, D.C. 20242

ATTN: Mr. William Hemphill (1)

Chief of Technical Support
Western Environmental Research Laboratories
Environmental Protection Agency
P.O. Box 15027
Las Vegas, Nevada 89114

ATTN: Mr. Leslie Dunn (1)

NASA/Langley Research
Mail Stop 470
Hampton, Virginia 23365

ATTN: Mr. William Howle (1)

U.S. Geological Survey
Branch of Regional Geophysics
Denver Federal Center, Building 25
Denver, Colorado 80225

ATTN: Mr. Kenneth Watson (1)

NAVOCEANO, Code 7001
Naval Research Laboratory
Washington, D.C. 20390

ATTN: Mr. J. W. Sherman, III (1)

U.S. Department of Agriculture
Administrator
Agricultural Stabilization and Conservation Service
Washington, D.C.

ATTN: Mr. Kenneth Frick (1)

Pacific Southwest Forest & Range Experiment Station
U.S. Forest Service
P.O. Box 245
Berkeley, California 94701

ATTN: Mr. R. C. Heller (1)

Pacific Southwest Forest & Range Experiment Station
U.S. Forest Service
P.O. Box 245
Berkeley, California 94701

ATTN: Dr. P. Weber (1)

NASA/Johnson Space Center
Mission Planning & Analysis Division
Houston, Texas 77058

ATTN: Mr. H. G. De Vezin/FM8 (1)

University of Texas at Dallas
Box 688
Richardson, Texas 75080

ATTN: Dr. Patrick L. Odell (1)

Department of Mathematics
University of Houston
Houston, Texas 77004

ATTN: Dr. Henry Decell (1)

Institute for Computer Services and Applications
Rice University
Houston, Texas 77001

ATTN: Dr. M. Stuart Lynn (1)

U.S. National Park Service
Western Regional Office
450 Golden Gate Avenue
San Francisco, California 94102

ATTN: Mr. M. Kolipinski (1)

U.S. Department of Agriculture
Statistical Reporting Service
Washington, D.C. 20250

ATTN: D. H. VonSteen/R. Allen (2)

U.S. Department of Agriculture
Statistical Reporting Service
Washington, D.C. 20250

ATTN: Mr. H. L. Trelogan, Administrator (1)

Department of Watershed Sciences
Colorado State University
Fort Collins, Colorado 80521

ATTN: Dr. James A. Smith (1)

Lockheed Electronics Co.
16811 El Camino Real
Houston, Texas 77058

ATTN: Mr. R. Tokerud (1)

TRW System Group
Space Park Drive
Houston, Texas 77058

ATTN: Dr. David Detchmendy (1)

IBM Corporation
1322 Space Park Drive
Houston, Texas 77058

ATTN: Dr. D. Ingram (1)

S&D—DIR
Marshall Space Flight Center
Huntsville, Alabama 35812

ATTN: Mr. Cecil Messer (1)

Code 168-427
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, California 91103

ATTN: Mr. Fred Billingsley (1)

NASA/Johnson Space Center
Technical Library Branch
Houston, Texas 77058

ATTN: Ms. Retha Shirkey/JM6 (4)

NASA Headquarters
Washington, D.C. 20546

ATTN: Mr. W. Stoney/ER (1)
ATTN: Mr. Leonard Jaffe/ER (1)
ATTN: Mr. M. Molloy/ERR (1)
ATTN: Mr. G. Thorley/ERR (1)

Ames Research Center
National Aeronautics and Space Administration
Moffett Field, California 94035

ATTN: Dr. I. Poppoff (1)

Goddard Space Flight Center
National Aeronautics and Space Administration
Greenbelt, Maryland 20771

ATTN: Mr. W. Nordberg, 620 (1)
ATTN: Mr. W. Alford, 563 (1)

Lewis Research Center
National Aeronautics and Space Administration
21000 Brookpark Road
Cleveland, Ohio 44135

ATTN: Dr. Herman Mark (1)

John F. Kennedy Space Center
National Aeronautics and Space Administration
Kennedy Space Center, Florida 32899

ATTN: Mr. S. Claybourne/FP (1)

NASA/Langley
Mail Stop 214
Hampton, Virginia 23665

ATTN: Mr. James L. Raper (1)